Nonlinear Time Series Analysis using Ordinal Networks with select applications in Biomedical Signal Processing

Michael Hugh McCullough
Bachelor of Electrical and Electronic Engineering (Hons); Bachelor of Music (Hons)
September 2017

School of Mathematics and Statistics

Supervisors:  Prof. Michael Small (Coordinating)
Dr. Thomas Stemler
Prof. Herbert H.C. Iu
Thesis Declaration

I, Michael Hugh McCullough, certify that:

This thesis has been substantially accomplished during enrolment in the degree.

This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

The works are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

This thesis contains published work and work prepared for publication, some of which has been co-authored.

Signature: [Redacted]

Date: 5/10/2017
Authorship Declaration: Co-authored Publications

This thesis contains the following 5 works that have been published:

Details of work 1:

Location in thesis:
Chapter 2

Student contribution to work:
Reviewing and summarising relevant literature; devising conceptual framework and hypothesis; writing code to generate numerical data and compute all results; interpreting results and establishing findings; formatting results and rendering figures; and writing manuscript for publication.

Details of work 2:

Location in thesis:
Chapter 3

Student contribution to work:
Reviewing and summarising relevant literature; devising conceptual framework and hypothesis; writing code to generate numerical data and compute results; devising method and writing code to analyse experimental ECG data; interpreting results and establishing findings; formatting results and rendering figures; and writing manuscript for publication.

Details of work 3:

Location in thesis:
Chapter 5 Sections 5.1, 5.2, and 5.3.

(Student 5.4 (unpublished) was completed entirely by the student).

Student contribution to work:
M. McCullough (the student) and K. Sakellariou contributed equally (50% each) to the following: Reviewing and summarising relevant literature; devising conceptual framework and hypothesis writing code to generate numerical data and compute all results; interpreting results and establishing findings; formatting results and rendering figures; and writing manuscript for publication.
Details of work 4:

Location in thesis:
Chapter 6

Student contribution to work:
M. McCullough (the student) contributed 75% and K. Sakellariou contributed 25% to the following: Reviewing and summarising relevant literature; devising conceptual framework and hypothesis; writing code to generate numerical data and compute all results; interpreting results and establishing findings; formatting results and rendering figures; and writing manuscript for publication.

Details of work 5:

Location in thesis:
This work was completed during the student’s PhD candidature but does not appear in this thesis. The research is a direct extension of the work presented in Chapter 6.

Student contribution to work:
M. McCullough (the student) contributed 25% and K. Sakellariou contributed 75% to the following: Reviewing and summarising relevant literature; devising conceptual framework and hypothesis; writing code to generate numerical data and compute all results; interpreting results and establishing findings; formatting results and rendering figures; and writing manuscript for publication.

Student signature:

Date: 5/10/2017

I, Michael Small, certify that the student statements regarding their contribution to each of the works listed above are correct.

Coordinating supervisor signature:

Date: 5/10/17.
Acknowledgments

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

The diode resonator dataset investigated in this thesis was kindly provided by Dr Thomas Jüngling.

I would like to express my sincere gratitude to my supervisors, Professor Michael Small, Dr Thomas Stemler and Professor Herbert Iu, for giving their time and expertise in support of my research, for being invaluable as mentors, and for providing me with numerous opportunities to grow my skills, experience and professional networks as an academic researcher.

I also wish to thank my contemporary, collaborator and good friend Konstantinos Sakellariou for his contributions to this work, and for our many lengthy discussions on mathematics, science and philosophy.

I would like to give very special thanks to my wife, Dee, for always being there to motivate and support me through this endeavour with her endless love and patience.

Finally, I thank my parents, Judith and Ian, for encouraging and supporting me through almost twenty-five years of formal education, and my brothers, Colin and Rowan, especially my older brother, Colin, whose brilliance and success before me has been inspirational and helped me believe in my own potential.
To my wife, Dee
Abstract

Over the past ten years, there have emerged numerous network-based methods for the analysis of nonlinear time series which provide novel and alternative means of extracting useful information from data measured from complex systems. The fundamental defining principle of such methods is that time series data are mapped or transformed to construct a network model, the structural properties of which can then be analysed using measures from the field of network science to infer dynamical properties of the system from which the time series has been observed. The underlying assumption is, therefore, that the network topology is consistently and measurably dependant on the time series dynamics. The subject matter of this thesis is the proposal, investigation and application of a new network-based method of nonlinear time series analysis which we term ordinal network analysis.

We define the procedure for constructing ordinal networks from univariate time series data in two stages. The data are first mapped to a sequence of discrete symbolic states called ordinal symbols by partitioning a delay embedding of the time series based on the amplitude rank of the elements within each embedded state vector. The finite symbol space is then bijectively mapped to a set of nodes in a network and directed edges are assigned between nodes based on temporal succession of states in the symbolic sequence. The ordinal network can therefore be interpreted as a Markov chain stochastic approximation of the time series dynamics.

The primary aim of this research is to investigate and establish the relationship between measures of ordinal network topology and time series dynamics for discrete-time sampled data from archetypal continuous autonomous chaotic systems. This aspect of the thesis is undertaken with respect to a range of simple network statistics and two newly proposed measures to characterise local transitional complexity and mixing rate respectively. Comparative numerical investigations with toy data and experimental chaotic circuit data show that the network measures reliably track the change in estimates of the largest Lyapunov exponent over a range of the bifurcation parameter, even in the presence of low levels of observational noise, demonstrating that network topology is measurably dependant on dynamics.

To test the potential of ordinal network analysis in practice, we apply our method in three separate investigations of biomedical time series. Firstly, we implement a multiscale framework for ordinal network analysis for discrimination between short time electrocardiogram recordings characterised by normal sinus rhythm, ventricular tachycardia and ventricular fibrillation respectively. Secondly, we use the same multiscale framework for the investigation of age-related effects in cardiac dynamics using interbeat interval time series from long time electrocardiograms. Thirdly, we apply the ordinal network method in a sliding window analysis of multivariate electroencephalogram time series to extract linearly separable feature vectors for epileptic seizure onset classification using binary linear support vector machines.

In addition, we investigate the ordinal network’s capacity to generate new time series with similar dynamical properties to the data used to construct the model. To do this we take constrained random walks on the network to regenerate new symbolic dynamics and use a data reassignment procedure to produce surrogate time series. We then compute invariant measures and recurrence properties to compare the original time series with the regenerated surrogates. Furthermore, we test the out-of-sample predictive properties of ordinal networks for low dimensional chaotic time series.

It has been established that the count of ordinal symbols which do not occur in a time series, called forbidden patterns, is an effective measure for the detection of determinism in noisy data. A very recent study has shown that this measure is also partially robust against the effects of irregular sampling. As a secondary investigation in this thesis, we extend said research with an emphasis on exploring the parameter space for the embedding dimension and find that the measure is more
robust to under-sampling and irregular sampling than previously reported. Using numerically generated toy data from discretely sampled continuous systems, we investigate the reliability of the relative proportion of ordinal patterns in periodic and chaotic time series for various degrees of under-sampling, random depletion of data, and timing jitter.
# Contents

## 1 Introduction

1.1 Nonlinear time series analysis ................................................................. 1
1.2 Network based methods ................................................................. 3
  1.2.1 Proximity networks ................................................................. 3
  1.2.2 Visibility graphs ................................................................. 5
  1.2.3 Transition networks ................................................................. 5
1.3 Ordinal time series analysis ................................................................. 6
1.4 Aims and general assumptions ................................................................. 7

## 2 Unweighted Ordinal Networks: Simple Stochastic Models for Capturing Complex Dynamics

2.1 Constructing unweighted ordinal networks from univariate time series ......................... 10
2.2 Basic measures for quantifying ordinal network structure ........................................... 12
  2.2.1 Network measures ................................................................. 12
  2.2.2 Quantifying network structure ..................................................... 14
2.3 Application: Chaotic voltage signal time series measured from a diode resonator circuit 22
2.4 Chapter summary ................................................................. 26

## 3 Weighted Ordinal Networks for Quantifying Transitional Complexity on Multiple Time Scales

3.1 Constructing weighted ordinal networks from univariate time series ........................... 28
3.2 Parameter selection ................................................................. 28
  3.2.1 Embedding dimension ................................................................. 29
  3.2.2 Embedding lag ................................................................. 31
3.3 Practical considerations ................................................................. 32
  3.3.1 Node aliasing ................................................................. 32
  3.3.2 Data with coarse quantisation ..................................................... 33
3.4 Quantifying the complexity of ordinal networks ........................................... 33
  3.4.1 Permutation entropy ................................................................. 33
  3.4.2 Measures of transitional complexity ................................................ 34
  3.4.3 Local and global node out-link entropy of ordinal networks ................. 35
### CONTENTS

3.5 Numerical investigation using the Rössler system ........................................... 36
3.6 Application: Electrocardiogram analysis .............................................................. 40
   3.6.1 Classifying cardiac dynamics using ordinal network entropy ............................. 40
   3.6.2 Heart rate variability analysis for different age groups using ordinal network entropy ................................................................. 46
3.7 Chapter summary ................................................................................................. 48

4 Characterising time series complexity via inverse normalised ordinal network diameter ........................................... 51

4.1 Markov chain mixing time and network diameter .................................................. 52
4.2 Ordinal network based estimates of system mixing time ......................................... 53
   4.2.1 Computing ordinal network diameter, the assumption of determinism, and the problem of self edges ................................................................. 53
   4.2.2 Improving estimates by re-weighting the ordinal network edges ......................... 54
   4.2.3 Obtaining irreducible ordinal networks in practice ............................................ 55
4.3 Numerical investigation ....................................................................................... 56
   4.3.1 The Rössler system ....................................................................................... 57
   4.3.2 The Muthuswamy-Chua system with quadratic nonlinearity (MC2) ................. 58
   4.3.3 The Muthuswamy-Chua system with 4th-order polynomial nonlinearity (MC4) .... 60
   4.3.4 The Chua circuit equations ............................................................................. 62
4.4 Application: Epileptic seizure onset detection and pre-detection from electroencephalograms .................................................. 64
   4.4.1 Data and background ................................................................................... 64
   4.4.2 Seizure onset detection ................................................................................ 66
   4.4.3 Identification of pre-seizure state .................................................................. 71
4.5 Chapter summary ................................................................................................. 74

5 Generating surrogate time series from ordinal network ............................................ 77

5.1 Generating time series from network models ......................................................... 77
5.2 Methodology ....................................................................................................... 79
   5.2.1 Preliminaries ............................................................................................... 79
   5.2.2 Random walks on ordinal networks ................................................................ 79
   5.2.3 Regenerating time series from random walks ............................................... 81
5.3 In-sample analysis of ordinal surrogate time series ................................................. 81
   5.3.1 Qualitative comparison ................................................................................ 82
   5.3.2 Order recurrence plots and recurrence quantification analysis ......................... 85
   5.3.3 Estimation of Lyapunov exponents .................................................................. 90
5.4 Out-of-sample prediction using ordinal networks .................................................... 92
   5.4.1 Prediction algorithm .................................................................................... 92
   5.4.2 Numerical investigation ............................................................................... 93
5.5 Chapter summary ................................................................................................. 94
6 Counting forbidden patterns in irregularly sampled time series
   6.1 Introduction ......................................................... 98
   6.2 Counting forbidden patterns ....................................... 98
   6.3 Numerical investigation ............................................. 100
      6.3.1 Under-sampling ................................................. 101
      6.3.2 Random depletion .............................................. 104
      6.3.3 Timing jitter ................................................ 106
   6.4 Chapter summary .................................................... 108

7 Summary and discussion .................................................. 111

Bibliography .................................................................. 117
List of symbols

\( t \) \quad \text{Continuous time}

\( \Delta t \) \quad \text{Sampling period}

\( n, k, l \) \quad \text{Discrete time indices}

\( x_n \) \quad \text{Scalar time series observation for discrete time index } n

\( \mathbf{x} \) \quad \text{Scalar time series comprising the ordered set of observations } x_n

\( N \) \quad \text{Length of time series}

\( m \) \quad \text{Embedding dimension}

\( \tau \) \quad \text{Embedding lag}

\( \mathbf{z}_n \) \quad \text{Embedded state vector}

\( s_n \) \quad \text{The ordinal symbol for discrete time index } n

\( s \) \quad \text{The symbolic dynamics of a given time series comprising the ordered set of symbols } s_n

\( S \) \quad \text{The unique set of ordinal symbols which occur in a given time series}

\( G \) \quad \text{Network or graph}

\( V \) \quad \text{The set of nodes in } G

\( \mathcal{E} \) \quad \text{The set of edges in } G

\( V \) \quad \text{Number of nodes in } G

\( A \) \quad \text{Adjacency matrix of } G

\( a_{i,j} \) \quad \text{Element of a stipulating a directed edge from node } i \text{ to } j

\( p_{i,j} \) \quad \text{Probability of a state transition from node } i \text{ to } j \ (\text{Equation 3.3})

\( k_{\text{out}} \) \quad \text{Node out-degree}

\( \bar{k}_{\text{out}} \) \quad \text{Mean node out-degree } (\text{Equation 2.2})

\( \psi \) \quad \text{Network diameter } (\text{Equation 2.3})

\( \chi \) \quad \text{Weighted network diameter } (\text{Equation 4.7})

\( \hat{\chi} \) \quad \text{Normalised weighted network diameter } (\text{Equation 4.9})
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{PE}$</td>
<td>Permutation entropy (Equation 3.2)</td>
</tr>
<tr>
<td>$h_{CPE}$</td>
<td>Conditional permutation entropy (Equation 3.6)</td>
</tr>
<tr>
<td>$h_{LNE}$</td>
<td>Local node out-link entropy (Equation 3.8)</td>
</tr>
<tr>
<td>$h_{GNE}$</td>
<td>Global node out-link entropy (Equation 3.9)</td>
</tr>
<tr>
<td>$\mathcal{P}_f$</td>
<td>Percentage of forbidden patterns in a time series (Equation 6.1)</td>
</tr>
<tr>
<td>$\lambda_k$</td>
<td>The $k$-th Lyapunov exponent in descending order</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Nonlinear time series analysis

Many of the challenges facing the modern scientific community concern the analysis or modelling of complex physical, biological or social systems. The complex nature of a system may arise, for example, from the interaction of a large number of simple connected components, or from the highly non-trivial time evolution of a low dimensional chaotic system, or some combination of the two. Whichever the case may be, the behaviour which manifests in complex real-world systems is often very difficult to understand or predict. Technological advances have, however, enabled us to measure, store and process large amounts of time series data from such systems with comparative ease. Nonlinear time series analysis is a field which encompasses a collection of tools and methods that have been designed to extract, quantify and visualise useful information from a time series under the assumption that the system under observation can be modelled as a nonlinear dynamical system.

The vast majority of these methods hinge on Takens’ embedding theorem (Takens [1981]). Informally, the theorem can be explained as follows. Consider the dynamical system given by the manifold $\mathcal{M} \subseteq \mathbb{R}^d$ and deterministic evolution operator $\Phi : \mathcal{M} \times \mathbf{V} \to \mathcal{M}$ such that $\Phi(v_n) = v_{n+1}$ where $v_n \in \mathcal{M}$ is the system state at the discrete time index $n$. Now define the measurement function $g$ which operates on $v_n$ to produce the univariate time series $g(v_n) = x_n$. Takens’ theorem states that the delay coordinate embedding $z_n = (x_n, x_{n+\tau}, x_{n+2\tau}, \ldots, x_{n+(m-1)\tau})$, with the embedding dimension $m$ and embedding lag $\tau$, is an object which is topologically equivalent to the original dynamical system for $m \geq 2d + 1$. While it can be easily understood that this equivalence is only guaranteed for an infinitely long and noise free time series, the practical implication is that a univariate time series which has been measured from a real world system can be embedded to reconstruct an attractor with topological invariants that are close to those of the true system.

Takens’ theorem has become the cornerstone of nonlinear time series analysis because it provides a means of investigating and modelling the phase space of systems that cannot be fully observed, which is almost always the case for real world systems. If one assumes that the delay reconstruction is an accurate model then a range of numerical methods can be applied to compute estimates of topological invariants or invariant-like quantities. Traditionally this would include the attractor dimension computed via the correlation integral (Grassberger and Procaccia [1983]), Kolmogorov-Sinai entropy (Sinai [1959, 1968]), and other related quantities.
or related complexity measures (Eckmann and Ruelle [1985]; Pincus [1991]), and the spectrum of Lyapunov exponents.

We now briefly address the latter because estimates of the largest Lyapunov exponent feature prominently in this thesis. The spectrum of Lyapunov exponents \( \{\lambda_1, ..., \lambda_m\} \) quantifies the rate of exponential divergence of nearby initial conditions, where the number of exponents is equal to the dimensionality of the system. The largest exponent \( \lambda_1 \) is directly related to a system’s time horizon for predictability and is most commonly used as a means of identifying chaotic dynamics from time series data because \( \lambda_1 > 0 \) implies chaos. Estimating the complete spectrum of exponents from time series is theoretically possible using the algorithm by Wolf et al. [1985] which first computes the rate of separation between the trajectory and a near neighbour in phase space to obtain \( \lambda_1 \), then the rate of growth of the surface defined by the trajectory and its two closest neighbours to obtain \( \lambda_1 + \lambda_2 \), and then the growth rate of hyper-surfaces in consecutively higher dimensions to determine the remaining exponents in the same manner. Alternatively, one can compute the complete spectrum of exponents from the tangent space as per the method’s proposed by Sano and Sawada [1985], Eckmann et al. [1986] and others. This, of course, necessitates the construction of a model of the dynamics from the delay coordinate embedding. Kantz [1994] notes that both Wolf’s algorithm and tangent space methods may produce erroneous results when applied to noisy time series and are sensitive to choice of embedding dimension: producing potentially misleading results when the dimension is too small; and spurious exponents that are difficult to identify when the dimension is too large. Significantly more robust methods are available if one is only interested in computing the largest exponent, namely those by Rosenstein et al. [1993] and Kantz [1994]. The fundamental premise of these methods is that one computes some distance metric \( q(z_{n+\Delta n}, z_{k+\Delta n}) \) between the state vector at time \( n \) and its nearest neighbour (or set of neighbours) at time(s) \( k \) for some range \( \Delta n \) in which one expects to observe exponential separation of trajectories\(^1\). The average distance \( \overline{q}(\Delta n) \) is computed over all \( z_n \) and the practitioner must then identify a clearly defined linear scaling region in the plot of \( \log(\overline{q}(\Delta n)) \) against \( \Delta n \) which is invariant to the embedding dimension \( m \). If such a scaling region exists, then this indicates exponential divergence of nearby initial conditions and is evidence of chaotic dynamics. One can then estimate \( \lambda_1 \) by taking the gradient of the scaling region. The need to identify a scaling region makes these methods far less prone to producing erroneous results that may arise from noisy data or from data which is not chaotic in the first place. However, the drawback is that the practitioner must have a thorough understanding of the method.

Broadly speaking, the assertion that we might be able to compute reliable estimates of topological invariants from a delay reconstruction of real world data is contingent upon: the selection of appropriate embedding parameters; the time series having sufficient length and sampling rate to capture the dynamics; that the data is of reasonably high quality (i.e. not too noisy); correct application of numerical methods; and, most importantly, the assumption that the data can be effectively modelled as a deterministic dynamical system. These issues have been addressed in considerable depth in the literature, including the recent comprehensive review by Bradley and Kantz [2015]. In practice there is always an element of uncertainty when estimating invariants so these measures are generally

---

\(^1\text{Rosenstein et al. [1993] uses Euclidean distance } |z_{n+\Delta n} - z_{k+\Delta n}| \text{ in the reconstructed phase space whereas Kantz [1994] takes the difference of the univariate time series points } |x_{n+\Delta n} - x_{k+\Delta n}|. \text{ The fact that the latter is a nonlinear projection of distance is accounted for when the value is averaged over all } z_n \text{ in the attractor.}\)
1.2 NETWORK BASED METHODS

used in conjunction with hypothesis testing using surrogate time series (Schreiber and Schmitz [2000]; Theiler et al. [1992]), statistical methods, or machine learning for relative classification or change point detection. Bradley and Kantz [2015] note that in this capacity, methods for estimating nonlinear invariants serve as a form of data compression, or a filter. This fact does not diminish the usefulness of these measures. If we suspect that a time series has been observed from a system with a strong nonlinear deterministic component in the dynamics, nonlinear measures are still likely to be more effective in quantifying the useful information contained in the data than linear methods. However, in applications where it is necessary to treat estimates of nonlinear invariants as statistics, it is reasonable to propose that there may exist other nonlinear measures or invariant-like quantities that are at least complimentary if not superior metrics — in either a statistical, numerical or practical capacity — for characterising specific features of the data.

1.2 Network based methods

Motivated in the pursuit of developing new tools for nonlinear time series analysis, Zhang and Small [2006] proposed the concept of transforming time series data into a network. Their method was to segment a pseudo-periodic time series into cycles which were then each bijectively mapped to a network node. Edges were subsequently assigned between nodes if the linear correlation between their corresponding cycles was sufficiently high with respect to a predetermined threshold. Results from the ensuing investigation uncovered links between the degree distribution of the network and distribution of unstable periodic orbits in the attractor that enabled distinction between noisy periodic time series and chaos. The concepts introduced in this seminal study have since been developed into a wide range of methods and frameworks that we shall refer to as network based time series analysis, as comprehensively reviewed by Donner et al. [2011] and more recently by Gao et al. [2016]. The defining principle of network based time series analysis is that data are mapped to a network given by a set of nodes and edges. Subsequent statistical analysis of the network’s topology is then used to infer and quantify properties of the time series dynamics. The critical attribute of such methods is the specific nature of the mapping procedure, which determines how information is encoded in the network and, therefore, how its structural properties can be interpreted in terms of the dynamics. In this regard, the majority of existing methods can be classified into one of three categories: proximity networks, visibility graphs, or transition networks.

1.2.1 Proximity networks

Proximity networks are constructed by defining a set of states from a time series, mapping each state to its own node, and assigning edges between nodes if the corresponding states are similar by some measure of closeness. Therefore, the networks proposed and investigated by Zhang and Small [2006] fall within this class of methods, where states are pseudo-periodic cycles in the data and closeness is determined by linear correlation of the cycles. It is more common, however, that states are simply embedded state vectors from a delay reconstruction and closeness is measured in terms of Euclidean distance in reconstructed phase space. An intuitive approach is to assign an edge between the nodes corresponding to a pair of embedded state vectors $z_n$ and $z_k$ if $|z_n - z_k| < \epsilon$, where $\epsilon$ is the connectivity
threshold parameter which can be tuned to control the edge density of the network. This is equivalent to reinterpreting a recurrence plot (Eckmann et al. [1987]) as a network adjacency matrix, and was first proposed by Marwan et al. [2009] who termed them recurrence networks. Xu et al. [2008] had previously proposed a similar concept based on a delay reconstruction called the $k$-nearest neighbour method. The critical distinction between $k$-nearest neighbour networks and recurrence networks is that the former assign edges between a state and it’s closest $k$ neighbours, rather than all neighbours within an $\epsilon$-ball as per the latter. This enforces connectivity independently of state density in phase space which may be beneficial in the investigation of highly heterogeneous attractors. Furthermore, the mean node degree and edge density of a $k$-nearest neighbour network are directly dependent on parameter $k$ which enables direct comparison between networks generated from different time series of the same length. We also note the approach proposed by Yang and Yang [2008] where nodes correspond to fixed-length segments of a time series (which can be considered as embedded state vectors under the assumption of determinism) and edges are assigned if data within a pair of segments has linear correlation greater than some threshold.

The recurrence network paradigm is arguably the most complete framework for network based time series analysis. The early work of Marwan et al. [2009] established relationships between the degree centrality and edge density of the network with local and global recurrence rate, respectively, which are traditional measures from recurrence quantification analysis (Marwan et al. [2007]). Additionally, the clustering coefficient and average shortest path length (mean geodesic) were shown to be effective for identifying change points in paleoclimate records. Donner et al. [2010] presented a detailed set of qualitative and analytical arguments in combination with numerical results to show that the spatial distribution of node properties including degree centrality, clustering coefficient, closeness centrality and betweenness centrality characterise attractor density and heterogeneity. They also suggest that clustering coefficient and similar measures may have potential applications in detecting chaotic saddles and unstable periodic orbits. Zou et al. [2012] uncovered power law degree distributions in recurrence networks generated from a variety of deterministic test systems and experimental data, and showed that the power law exponent is directly related to the invariant density of the underlying systems. Donner et al. [2011] and Xiang et al. [2012] have also established relationships between network topology and attractor heterogeneity for $k$-nearest neighbour networks, with specific emphasis on the structure of unstable periodic orbits. Perhaps more significant, however, was the finding by Xu et al. [2008] (revisited in Xiang et al. [2012]) that distribution of subgraphs of small fixed size in $k$-nearest neighbour networks can be used to discriminate between chaotic and noisy periodic dynamics. The authors postulated that the edge configuration in these subgraphs is related to the local dimension of the attractor and, therefore, characterises the dynamics.

Proximity networks are useful as a means of encoding and quantifying attractor topology. Donner et al. [2010] points out that the primary limitation of these methods is that they are invariant to temporal reordering of embedded state vectors and, therefore, do not contain information about dynamical evolution of trajectories. In addition, it is necessary to select parameters for the delay coordinate embedding and a third parameter as a connectivity threshold. Furthermore, it is generally advisable to specify a fourth parameter which defines a Theiler window to ensure that temporally adjacent states are not connected (Donner et al. [2011]). While this is a relatively small parameter
1.2. NETWORK BASED METHODS

1.2. NETWORK BASED METHODS

1.2. NETWORK BASED METHODS

1.2. NETWORK BASED METHODS

1.2. NETWORK BASED METHODS

space, the topology of a proximity network can be highly dependant on the selection of the connectivity parameter (Donner et al. [2010]), and fewer method parameters are always preferable.

1.2.2 Visibility graphs

Visibility graphs were first proposed by Lacasa et al. [2008]. These are computationally simple procedures that map each point in the time series to a node and then connect the nodes based on the convexity of successive observations. Aside from several extensions proposed for specific tasks (Ahmadlou and Adeli [2012]; Bezsudnov and Snarskii [2014]), visibility graph methods are parameter free. Research has shown links between scale free networks and fractal time series (Lacasa et al. [2008]), and that the degree distributions of visibility graphs generated from fractal Brownian motion are a function of the Hurst exponent (Lacasa et al. [2009]). Relationships between visibility graph topology and properties of deterministic dynamical systems are currently unexplored for the most part. One notable exception is the work of Donges et al. [2013] where visibility graphs were used to test for symmetry in time series and therefore detect evidence of nonlinear deterministic dynamics. The premise of this study was to separately compute local node properties for edges that connect in forward time from those which connect in reverse time. If the time series used to construct the network is irreversible, then one would expect that the forward and reverse time node statistics will be drawn from different distributions. The authors exploited this characteristic to test for nonlinearity in various toy data and human electroencephalograms without the need for surrogate data.

1.2.3 Transition networks

Consider partitioning a univariate time series or delay coordinate embedding into a set of disjoint subspaces which span the data. Each datum or embedded state vector is then mapped to one of a set of discrete symbols corresponding to the elements of the partition, resulting in symbolic sequence that describes the time series trajectory. Modelling dynamics from a continuous space as a discrete set of symbols in this manner is known as symbolic dynamics (Marcus and Williams [2008]). A transition network is constructed by treating the unique set of symbols as a set of nodes and then assigning directed links between nodes based on immediate temporal succession in the symbolic dynamics. Therefore, the network models the symbolic trajectory as a 1-step shift of finite type. Alternatively, the network can be interpreted probabilistically as a first-order Markov chain. Transition networks are primarily useful as means to model and investigate the temporal evolution of systems.

While symbolic dynamics and Markov chain theory are both long-established fields, the concept of building these kinds of models from time series data and treating them as networks has seen renewed interest in recent years. Notable works include Padberg et al. [2009] where subgraph expansion rates were computed to detect transport barriers in flows, and Froyland and Padberg-Gehle [2012] in which the authors defined a measure based on the entropy of successive states to quantify local nonlinear stretching. Furthermore, Ros et al. [2013] show relationships between the predictability of dynamical systems and traditional network measures including node centrality and network diameter. A uniform grid was applied to partition phase space in the aforementioned studies. Campanharo et al. [2011] proposed an alternative approach whereby the discretely sampled univariate time series was partitioned by quantiles. Either approach will provide some degree of robustness to noise but inherently will
also discard information about oscillations of small amplitude. Ideally, one seeks what is known as a
generating partition which, by definition, will yield symbolic dynamics that are topologically conjugate
to phase space and enable the computation of Kolmogorov-Sinai entropy if the time series is noise-free.
However, even if one has access to long and noise free time series, estimating generating partitions is a
highly non-trivial exercise and seldom successful in practice. The question then remains as to how best
to partition a time series or delay coordinate embedding to construct a useful model of the dynamics.

1.3 Ordinal time series analysis

An ordinal partition defines a map from \( m \)-dimensional state space to the set of discrete symbols
given by the rank ordering or, equivalently, the amplitude rank of the elements of a state vector
(see Figure 1.1). This map was first conceived by Bandt and Pompe [2002] as part of a method to
estimate a measure of time series complexity known as permutation entropy, which is computed by
taking the Shannon entropy of the empirical probability distribution of the ordinal symbolic dynamics.
Permutation entropy and other forms of time series analysis based on ordinal symbolic dynamics
have been collectively termed ordinal analysis by Amigó et al. [2014]. These methods have found
numerous applications in biomedical signal analysis including heart rate variability analysis from
electrocardiograms (Bian et al. [2012]; Graff et al. [2013]; Makowiec et al. [2015]; Parlitz et al. [2012];
Zunino et al. [2015]), and electroencephalogram analysis for seizure onset detection, sleep staging, and
anaesthesiology (Groth [2005]; Keller et al. [2014]; Li et al. [2010]; Zunino et al. [2015]). We refer
readers to the recent papers by Zanin et al. [2012] and Amigó et al. [2014] for a thorough review of
ordinal analysis in biomedical applications.

![Figure 1.1: The set of all possible ordinal symbols of length \( m = 3 \) represented as tuples by rank ordering. Data points are ranked by amplitude in descending order.](image)

Generally speaking, the theory underpinning the findings of the aforementioned applications is thus
far incomplete, however, several key properties of ordinal analysis give merit to continuing work in the
field. Firstly, Keller and Sinn [2010] and Amigó and Keller [2013] have shown that an ordinal partition
has the generating property under certain conditions. Secondly, Amigó et al. [2007] demonstrated with
respect to discrete maps that the distribution of ordinal symbols present in time series is governed
by the dynamical properties of the system. Various methods and metrics for ordinal analysis have
been proposed based on these properties including: Bandt and Pompe’s [2002] original permutation
entropy measure and a range of derivative measures based on the statistical complexity of the empirical
probability distribution of ordinal symbols (Lange et al. [2013]; Soriano et al. [2011]; Zunino et al. [2010, 2012]); ordinal analysis for change point detection and data classification (Amigó et al. [2014]; Bandt and Shiha [2007]; Graff et al. [2013]; Makowiec et al. [2015]; Parlitz et al. [2012]); detection of determinism in noisy time series via forbidden pattern analysis (Amigó et al. [2007, 2010]; Rosso et al. [2012]); detection of synchronisation by ordinal recurrence plots (Amigó et al. [2014]; Groth [2005]), joint permutation entropy (Arroyo et al. [2013]) and transcripts (Aschenbrenner et al. [2013]) respectively; and forecasting with local linear models constructed over ordinal partitions (Hirata and Aihara [2012]; Paucar Bravo et al. [2013]). Furthermore, in a very recent and noteworthy study, Politi [2017] proposed an extension to Bandt and Pompe’s [2002] original method for computing permutation entropy that incorporated the probability distribution of ordinal symbols and the width of cylinder sets containing all time series trajectories corresponding to each respective symbol. The resulting measure was shown to converge on the Kolmogorov-Sinai entropy far more rapidly than permutation entropy and can also be used as a tool to estimate fractal dimension.

Ordinal approaches to time series analysis also present a number of practical advantages. They are inherently robust to some degree of observational noise because the symbolic mapping depends on the relative value of each datum with respect to its temporal neighbours (Amigó et al. [2010]; Bandt and Pompe [2002]). For the same reason, ordinal methods are invariant to order-isomorphisms (Amigó et al. [2007]) which is a useful property when dealing with mildly non-stationary data. Finally, generating the ordinal symbolic dynamics from a time series is generally accepted as being computationally efficient (Amigó et al. [2014]; Keller et al. [2014]; Monetti et al. [2013]; Unakafova and Keller [2013]).

1.4 Aims and general assumptions

In a preliminary study, Small [2013] recently proposed to construct transition networks based on an ordinal partition of reconstructed phase space, henceforth referred to as ordinal networks. The primary aim of this thesis is to investigate and develop a practically applicable framework for the analysis of complex systems from time series data using ordinal networks. Our hypothesis is that the topology of an ordinal network will be consistently and measurably dependant on dynamical properties of the time series from which it has been constructed. Specifically, we study the relationship between various local and global measures of network topology and properties of the time series which characterise some aspect to the temporal evolution of the system from which the time series has been observed. The methods and measures we propose are not only tested against numerically generated data from various low dimensional chaotic systems, but also applied for classification and change-point detection in several well studied biomedical datasets.

The scope of this work is limited to discretely sampled time series from continuous systems. We proceed with this thesis under two critical assumptions. Firstly, the time series data investigated henceforth are assumed to be stationary as per the definition given by Kantz and Schreiber [2004]. That is, a time series is stationary if the joint probability of observing some state $v_k$ given prior observation of the state $v_n$ is time independent, where $n, k$ are time indices such that $n < k$. This condition implies not only that the process governing the dynamics is time independent but also that the time series is sufficiently long such that it contains enough information to describe intermittent phenomena.
Secondly, we assume that the ergodic hypothesis applies to the systems which we investigate (Eckmann and Ruelle [1985]).

The remainder of this thesis is structured as follows. In Chapter 2 we define and investigate the simplest realisation of an ordinal network, where edges are directed but unweighted, using fundamental measures of networks topology. In Chapter 3 we extend our definition of the ordinal network such that edges are weighted based on the frequency of state transitions, and estimate local transitional complexity by computing the entropy over the out-connected edges of each node. We then use this measure in a newly proposed scheme for multiscale heart rate variability analysis. In Chapter 4 we investigate the relationship between network diameter and mixing rate. We propose the modification of edge weights to encode the average rate of local temporal evolution for each node and compute normalised network diameter as a metric for complexity. This new measure is then applied to multivariate electroencephalogram records from individuals with epilepsy to classify seizure onset. In Chapter 5 we investigate the extent to which ordinal network models can generate new time series with similar dynamics to the data from which the model was constructed. Furthermore, we propose and test a method for time series prediction using ordinal networks. The thesis concludes in Chapter 6 with an investigation of the robustness of the count of forbidden patterns — a statistic of the distribution of ordinal symbols enumerated from a time series — when the data is non-uniformly sampled.
Chapter 2

Unweighted Ordinal Networks: Simple Stochastic Models for Capturing Complex Dynamics

ABSTRACT

In Chapter 1 we reviewed a wide array of existing network based methods for addressing the problem of how to analyse univariate time series data measured from complex nonlinear systems. Of these methods, those which have been most thoroughly researched and established as frameworks are generally based on the concept of constructing networks where nodes correspond to states and edges are allocated based on measures of closeness or similarity between states. The resulting models are well suited for the study of topological properties of reconstructed phase space, but less useful for investigating the temporal relationships between states.

We concern ourselves with the latter and, therefore, one logical approach is to construct networks as Markov chains which constitute stochastic models that approximate the deterministic dynamics underlying a set of time series observations. The question which follows is, how might we best define the states in our model? Our method, and one novel aspect of this thesis, is that we define states by applying an ordinal partition to reconstructed phase space. As explained in the preceding chapter, an ordinal partition generates symbolic dynamics with an alphabet that comprises order patterns enumerated from the set of embedding vectors, and has several notable advantages including computational efficiency and parametric simplicity.

What follows in this chapter is our definition and investigation of the simplest form of a two parameter ordinal network, whereby edges are directed but unweighted. We examine the relationship between the system dynamics and ordinal network structure to show that basic network measures can be used to infer properties of time series. This is achieved via the numerical investigation of discrete sampled data from a continuous low dimensional chaotic system, and with application to voltage signal time series from an experimental diode resonator circuit.
2.1 Constructing unweighted ordinal networks from univariate time series

Herein we formally define the method for constructing unweighted ordinal networks from univariate time series data. This process essentially comprises two parts. The first is to map the time series data to the set of ordinal symbols as per the method of Bandt and Pompe [2002], and the second is to construct a network model from the symbolic dynamics. Readers should refer to Figure 2.1 for an illustrated outline of the procedure in the forthcoming definition.

Consider a series of uniformly sampled observations \( x = \{x_n\}_{n=1}^{N} \). To generate an ordinal network corresponding to \( x \) the data is first partitioned based on ordinal patterns of length \( m \) and time lag \( \tau \). To do this we take the series of embedding vectors \( \{z_n\}_{n=1}^{N-(m-1)\tau} \) where \( z_n = (x_n, x_{n+\tau}, x_{n+2\tau}, ..., x_{n+(m-1)\tau}) \). Each vector \( z_n \) is then mapped to a symbol \( s_n = (\pi_1, \pi_2, ..., \pi_m) \) where \( \pi_k \in \{1, 2, ..., m\} \) and \( \pi_k \neq \pi_l \iff k \neq l \) such that \( \pi_k < \pi_l \iff x_k > x_l \), for all \( x_k, x_l \in z_n \). In the case where two elements of \( z_n \) are equal in value, rank is assigned based on order of appearance in the vector, that is to say \( \pi_k < \pi_l \iff k < l \) for \( x_k = x_l \). We call \( S \) the unique set of all the ordinal symbols enumerated from the time series. Thus far we have applied an ordinal partition to \( \{x_n\}_{n=1}^{N} \) and can write the symbolic dynamics \( s = \{s_n\}_{n=1}^{N-(m-1)\tau} \) for \( s_n \in S \) corresponding to the time series.

The next step in the method is to construct a network \( G = (\mathcal{V}, \mathcal{E}) \) where \( \mathcal{V} = \{1, 2, ..., V\} \) is the set of nodes for \( V = |\mathcal{V}| \) and \( \mathcal{E} \) is the set of edges. We do this by applying the bijective operator \( \Gamma : S \rightarrow \mathcal{V} \), hence each node corresponds to an ordinal symbol from \( S \). We represent this network by the \( V \times V \) adjacency matrix \( A \) where the binary scalar element \( a_{i,j} \in \{0, 1\} \) stipulates whether or not an edge exists between node \( i \) and node \( j \). Edges in an unweighted ordinal network are assigned based on the condition:

\[
a_{i,j} = \begin{cases} 
1 & \text{if } \exists n \in \{1, 2, ..., (N - (m - 1)\tau)\} : (\Gamma(s_n), \Gamma(s_{n+1})) = (i, j) \\
0 & \text{for all } s_n \in S \text{ and all } i, j \in \mathcal{V} \text{ where } i \neq j ,
\end{cases}
\]

(2.1)

which is to say we place a directed edge from node \( i \) to node \( j \) if and only if there exists at least one instance where the symbol corresponding to \( i \) is followed immediately by the symbol corresponding to \( j \) in the symbolic dynamics \( s \). The resulting ordinal network is therefore a simple Markov chain approximation of the time series dynamics where all state transitions in the set of possible transitions for a given node are equi-probable regardless of their frequency in \( s \). It is also important to note that equation 2.1 implies that the model excludes the possibility for self edges (i.e. \( a_{i,i} = 1 \)). We specify this exclusion because a self edge in an unweighted ordinal network has one of two trivial implications under the assumption of determinism: the node corresponds to a fixed point in phase space or; the time taken to traverse the region of phase space corresponding to the node is greater than the interval at which the continuous system has been discretely sampled. With regard to the latter, the information encoded in a self edge can be useful when the ordinal network edges are weighted based on the frequency of transitions as enumerated from \( s \), because the weight of a self edge tells us about the rate of evolution of a system. This is a concept which we shall explore in Chapters 3 and 4. However, in the unweighted...
2.1. CONSTRUCTING UNWEIGHTED ORDINAL NETWORKS FROM UNIVARIATE TIME SERIES

Figure 2.1: An illustration of the process for transforming a univariate time series into an ordinal network: (a) Take the first embedding vector with lag $\tau$ and dimension $m$, (b) find its ordinal pattern which is the amplitude rank of its elements, (c) repeat the process for all embedding vectors from the time series, and (d) define a network $A$ with a node for each unique ordinal pattern, and edges assigned between nodes based on the temporal succession of ordinal patterns in the time series.

case which we consider henceforth, self edges would serve only to obscure the transitional structure encoded in the network and distort measures used to quantify complexity or uncertainty.

An unweighted ordinal network is fully described by its input time series and the two free parameters of the method: the embedding lag $\tau$ and the embedding dimension $m$. Ordinal symbolic methods were first conceived by Bandt and Pompe [2002] without a lag variable and the elegance of a single parameter approach to nonlinear time series analysis is undeniable. Alternatively, using time lagged ordinal symbols, as first proposed by Cao et al. [2004], enables the symbolic dynamics to encode information about trajectories on longer time scales whilst maintaining a smaller embedding dimension. It was shown by Zunino et al. [2010] that the lag parameter also appears to maximise the encoding of dynamical information for specific time scales when it is set appropriately, as shall be discussed in Chapter 3. In this chapter, however, we proceed under the simple assumption that $\tau$ should be selected to provide a suitable reconstruction of phase space for the time series under investigation. Therefore, the selection of $\tau$ is based on traditional criteria used for time delay embedding as detailed in the works of Small [2005] and Kantz and Schreiber [2004] respectively. In this chapter we have used the first zero of the autocorrelation of the time series. Other methods such as mutual information could foreseeably be used where they provide a better reconstruction for the particular data in question.

The embedding dimension $m$ controls how finely the time series is partitioned. Larger $m$ results in a finer partition which may afford a more accurate network model of the dynamics. The number of possible symbolic states in an ordinal partition is bounded from above by $m!$ and, hence, the size of an ordinal network $V$ grows rapidly as $m$ increases. This places a practical limitation on the choice for $m$ if one is attempting to compute certain measures on the network. For example, the fastest algorithms for computing shortest paths and enumerating loop distributions are, respectively: the
algorithm by Dijkstra [1959] implemented using a Fibonacci heap as per the work of Fredman and Tarjan [1987] which is bounded by $O(|E| + V \log(V))$ operations; and the algorithm by Johnson [1975] which is bounded by $O((V + |E|)(C + 1))$ operations where $C$ is the number of distinct cycles in the network. It becomes impractical to execute either of these computations when $V$ is large. A further, more important practical limitation arises from the requirement that the data length $N$ must be sufficiently long to provide good sampling over the partitioned reconstructed phase space, which shall be discussed in Chapter 3. Thorough discussion of the embedding dimension and how it controls the topology of the ordinal partition is also included in Chapter 3.

As a final comment on the method of constructing unweighted ordinal networks, observe that we have used fully overlapping embedding vectors in the map between the time series $x$ and the symbolic dynamics $s$. The use of non-overlapping vectors for ordinal analysis of time series was first proposed in Cao et al. [2004] and subsequently implemented in several studies including Amigó et al. [2010] and Small [2013]. Given that all forms of ordinal analysis depend on the assumption of ergodicity and stationarity, the properties of an ordinal symbolic dynamics for an infinite time series will be invariant to the choice of time index $k$ of the first embedding vector in the series of $\{z_n\}_{n=k}^{\infty}$. Hence, in the case of ordinal analysis techniques based on the statistical distribution of symbols, as reviewed in Chapter 1, the use of non-overlapping embedding vectors does nothing more than discard vast amounts of data. For example, if $m = 2$, a non-overlapping ordinal map will discard half of all the observable patterns in the data, and more will be discarded as $m$ increases. If a non-overlapping ordinal map was used for the creation of ordinal networks, the result could potentially be even worse because the data is not only being discarded but the resulting downsampling of the time series will cause aliasing in the symbolic dynamics that can significantly affect both local and global statistics on the network, as will be discussed in Chapter 3. Therefore we recommend the use of fully overlapping embedding vectors unless more sparsely sampled symbolic dynamics are sufficient to yield a result in specific applications where computational resources are limited.

## 2.2 Basic measures for quantifying ordinal network structure

In this section we define and discuss several simple network measures and how they can be interpreted in the context of ordinal network analysis for quantifying properties of time series data.

### 2.2.1 Network measures

**Network size**

The network size, or number of nodes in the network $V$, is one of the simplest quantifiable properties of any network. By definition, $V = |S|$ which is the number of distinct symbolic states visited by the trajectory $\{z_n\}_{n=1}^{N-(m-1)r}$ in reconstructed phase space. Therefore, for time series of fixed length $N$, the network size $V$ can be considered a measure of global topological complexity, albeit a coarse one, on the basis that a complex system is more likely to visit a greater number states in a fixed interval of time than a system that is less complex. Furthermore, given that each node corresponds to a disjoint region of phase space and assuming that reconstructed phase space is bounded, the quantity $V/m!$ describes the extent to which the reconstructed phase space has been filled. Alternatively, one can consider the
complement of this value which describes the fraction of possible ordinal patterns of length \(m\) that do not occur in the time series. It was established by Amigó et al. [2007] that the existence of such a set of forbidden patterns is a fundamental property of dynamical systems. We note here that the global complexity of the ordinal symbolic dynamics \(s\) can be quantified more completely by computing the permutation entropy of the time series (Bandt and Pompe [2002]), which is the Shannon entropy of the empirical probability distribution of \(s\). Permutation entropy shall be discussed in Chapter 3 in relation to weighted ordinal networks.

**Mean and variance of node out-degree**

The out-degree \(k_{\text{out}}\) of a node \(i\) quantifies the local transitional complexity of the network model at \(i\) with respect to the forward time evolution of the system. Conversely, the in-degree \(k_{\text{in}}\) quantifies local complexity with respect to the reverse-time evolution of the system. Taking the arithmetic mean gives a global measure of transitional complexity:

\[
\bar{k}_{\text{out}} = \frac{1}{V} \sum_{i=1}^{V} k_{\text{out}}(i) . \tag{2.2}
\]

It follows that the variance of the out-degree \(\text{Var}(k_{\text{out}})\) will be related to the heterogeneity of the model and the reconstructed phase space, thereby quantifying a different facet of complexity. Intuitively, \(k_{\text{out}} > 1\) implies the divergence of initial conditions from the neighbourhood in phase space defined by the ordinal partition element corresponding to node \(i\). However, because \(k_{\text{out}}\) is a measure on a stochastic model approximating the underlying system, nothing can be ascertained from \(\bar{k}_{\text{out}}\) or \(\text{Var}(k_{\text{out}})\) as to whether the mechanism driving the divergence is deterministic or stochastic. Furthermore, \(k_{\text{out}} > 1\) can arise in an ordinal network constructed from a strictly periodic time series due to aliasing effects, as will be discussed in Section 3.3.1.

In Chapter 3 we define a more accurate measure of local complexity by computing the entropy of the distribution transitional probabilities at each node in a weighted ordinal network. However, we will demonstrate in this chapter that the simple measure \(\bar{k}_{\text{out}}\) is nonetheless highly sensitive to changes in dynamics.

**Network diameter**

The diameter of an unweighted ordinal network is defined as:

\[
\psi = \sup \{g_{i,j}\} \text{ for all } \{i, j : i \neq j\} \in V , \tag{2.3}
\]

where \(g_{i,j}\) is the shortest path from node \(i\) to node \(j\) in the ordinal network. It is a global measure that is highly sensitive to any change in the dynamics which creates shortcuts in the network structure such as short bursts of intermittent chaotic dynamics. In Chapter 4 we investigate a refinement of this simple measure using weighted ordinal networks, where classical Markov chain theory allows us to define a new measure related to the rate at which a system mixes.
2.2.2 Quantifying network structure

To investigate the relationship between the dynamical properties of a time series and the structure of its corresponding ordinal network, we generate discrete sampled time series using the widely studied low dimensional chaotic system devised by Rössler [1976]:

\[
\begin{align*}
\frac{dx}{dt} &= -y - z, \\
\frac{dy}{dt} &= x + \alpha y, \\
\frac{dz}{dt} &= \beta + z(x - \gamma),
\end{align*}
\]

where \(\beta = 2\) and \(\gamma = 4\). We solve these equations numerically using a fourth-fifth order Runge-Kutta algorithm with randomised initial conditions \(x_0, y_0, z_0 \in [0, 1]\) for 1201 evenly spaced values of the bifurcation parameter \(\alpha \in [0.37, 0.43]\). The systems are solved for \(t \in [0, 4000]\) and discretely sampled at \(\Delta t = 0.2\). The first 10000 points corresponding to \(t \in [0, 2000]\) are discarded to remove transients. In the investigations to follow we construct ordinal networks using the resulting \(N = 10000\) point long \(x\)-component time series, which all have approximately 30 points per period-1 equivalent cycle. The first zero of the autocorrelation function was consistently found to occur at \(\tau \approx 9\) for the full domain of \(\alpha\) considered. We use this parameter value for constructing the networks. Henceforth, all network visualisations have been produced using a spring electrical embedding algorithm as implemented in Wolfram Mathematica 11. This algorithm positions vertices at an equilibrium point in 3-dimensional space as if the network were a physical system where nodes experience an attractive spring force proportional to their separation in Euclidian space if they are connected, and a repulsive force that is inversely proportional to Euclidian separation regardless of connectivity.

**Network structure versus embedding dimension**

Figure 2.2 is a simple qualitative illustration of the effect of the embedding dimension \(m\) on the structure of ordinal networks constructed from a period-2 time series. The complete cycle is approximately 62 samples long and each sub-cycle is half of this period. Consider that an ordinal symbol spans \((1 + (m - 1)\tau)\) samples. Therefore, when \(m = 6\) and \(\tau = 9\) (Figure 2.2a), the symbol spans 46 samples, which is significantly less than the complete period-2 cycle. It follows that each ordinal symbol may not contain sufficient information about the trajectories to distinguish between the first and second sub-cycles. This is a degeneracy in the ordinal symbolic map and results in a network where the paths corresponding to distinct trajectories are folded onto each other. Increasing the embedding dimension to \(m = 8\) creates a finer partition where each symbol spans 64 samples or just over one complete cycle (Figure 2.2b). It can be observed that the network has almost unfolded into a single periodic path but several degeneracies are still manifest as a knot in the centre of the structure. When \(m \geq 10\) such that the ordinal symbol spans 82 samples (Figures 2.2c and 2.2d), the complete period-2 trajectory is unfolded over the ordinal network as a ring structure.

The case for chaotic time series is similar, as shown in Figure 2.3. As \(m\) increases, the network unfolds into something that resembles the chaotic Rössler attractor, and for \(m = 10\) one can observe regions of the network corresponding to the stretching and folding of the chaotic trajectories respectively.
2.2. BASIC MEASURES FOR QUANTIFYING ORDINAL NETWORK
STRUCTURE

Figure 2.2: Four ordinal networks constructed from the same $x$-component time series generated by the Rössler system in a period-2 regime with bifurcation parameter $\alpha = 0.37$. The networks are constructed with embedding dimensions (a) $m = 6$, (b) $m = 8$, (c) $m = 10$, and (d) $m = 12$ respectively (increasing from left to right). The embedding lag is set at $\tau = 9$.

Figure 2.3: Four ordinal networks constructed from the same $x$-component time series generated by the Rössler system in fully developed chaos with bifurcation parameter $\alpha = 0.4$. The networks are constructed with embedding dimensions (a) $m = 6$, (b) $m = 8$, (c) $m = 10$, and (d) $m = 12$ respectively (increasing from left to right). The embedding lag is set at $\tau = 9$.

However, unlike the periodic case, it is more difficult to identify a value of $m$ which results in a network that constitutes a suitably useful model of the complex dynamics. Ordinal partitions have the generating property (Amigó and Keller [2013]; Keller and Sinn [2010]) and, as such, ordinal symbolic dynamics can uniquely describe an arbitrary trajectory, distinct from any other trajectory given $N \to \infty$, sufficiently large $m$, and under the condition that the trajectories are not an order isomorphic pair. In practice, such a model of a chaotic system would be trivial because it would require $m : V = (N - (m - 1)\tau)$ to completely describe an arbitrary finite non-repeating trajectory on the attracting set. An arguably better objective is to identify a value of $m$ large enough such that the model captures the facets complexity that are of interest (i.e. divergence of trajectories, chaotic mixing etc.), yet small enough such that the model description is much smaller than the data (i.e. $V \ll N$) and is therefore a useful compressed representation of the dynamics. For example, the network in Figure 2.3d has $V = 1813$ which is approaching the same order of magnitude as the length of the time series. The sampling rate of the data must also be considered when using a fine partition. If the data has been sampled at an insufficient frequency with respect to the relative size of the partition elements, edges can skip over states which would otherwise be visited by the true continuous trajectory and falsely encode additional complexity into the model (see discussions pertaining to node aliasing in Chapter 3). The combination of finiteness of data with respect to $m$ and insufficient sampling rate is likely the reason why the structure of the ordinal network for $m = 12$ in Figure 2.3d appears qualitatively less like a Rössler attractor than those constructed with smaller $m$. 
Figure 2.4: (a) The number of nodes in the network $V$, (b) mean node out-degree $\bar{k}_{out}$, (c) variance of node out-degree $\text{Var}(k_{out})$, and (d) the network diameter that has been normalised with respect the size of the network $\psi/V$ plotted against the embedding dimension $m$. Each curve corresponds to ordinal networks constructed from one of five $x$-component time series generated by the Rössler system in different dynamical regimes with bifurcation parameters $\alpha = 0.37$ (period-2), $\alpha = 0.41$ (period-3), $\alpha = 0.39$ (multiband chaos), $\alpha = 0.4$ (broadband chaos) and $\alpha = 0.42$ (broadband chaos) respectively. The embedding lag is set at $\tau = 9$. 
Given that our primary objective is to quantify network structure to infer the underlying time series dynamics, we compute the simple network measures defined in Section 2.2.1 against $m$ to investigate their dependence on the embedding dimension. Results for a selection of different time series are shown in Figure 2.4. Interesting features can be observed in all of these plots for $m \approx 8$. Ordinal networks constructed from chaotic time series grow far more rapidly than those constructed from periodic data when $m > 8$. The mean degree $\bar{k}_{out}$ for periodic time series grows and then saturates for this same embedding dimension. This implies that the network structure is stabilising despite slow growth in the network size which agrees with the preceding qualitative analysis in Figure 2.2. In Figure 2.4d we normalise $\psi$ with respect to the network size $V$ to enable comparison over the domain of $m$. For $m > 8$, the normalised ordinal network diameter for chaotic time series appear to converge on a minimum value. Furthermore, clear maxima can be observed for $m \approx 8$ in the measures $\bar{k}_{out}$ and $\text{Var}(k_{out})$ for networks constructed from chaotic time series. For larger $m$ these measures decrease rapidly. These maxima imply that networks constructed with $m \approx 8$ have the greatest local complexity and heterogeneity which, in turn, suggests that this parameter value maximises the amount of information in the ordinal network model for the data under investigation. Finally, the measures $\bar{k}_{out}$, $\text{Var}(k_{out})$ and $\psi/V$ show maximum discrimination between the periodic and chaotic time series dynamics for $m \approx 8$. It should be stressed that this analysis does not amount to a method for the selection of an optimum embedding dimension. However, given that local network complexity appears to be maximised at $m = 8$, and that aspects of the network structure for both periodic and chaotic time series appear to stabilise for larger embedding dimensions, we elect to proceed with our investigations using this parameter value.

Network structure versus length of the time series

Experienced practitioners in the field of applied nonlinear time series analysis will know all too well that real world time series data are often shorter in duration than we would like. To test the sensitivity of ordinal network structure with respect to the data length $N$ we generated five additional Rössler time series from different dynamical regimes each with $N = 100000$ (after the removal of transients) and constructed networks for increasing $N \in [100, 100000]$. Figure 2.5 shows simple network measures computed with respect to $N$. Network measures converge to stable values within 50-100 cycles (i.e. $N < 5000$) when the dynamics are periodic, which is to be expected given that the dynamics are fully described by a single cycle of the continuous limit cycle trajectory. Convergence would likely be faster for smaller $\Delta t$. Chaotic systems can take many cycles to fill out their attractor sufficiently such that a univariate time series contains a good enough representation of the underlying dynamics for data driven modelling methods to be implemented effectively. Network measures computed for chaotic time series generally converge to stable values for $N \approx 20000$. However, there is still good discrimination between the time series dynamics for as low as $N \approx 5000$. 

17
Figure 2.5: (a) The number of nodes in the network $V$, (b) mean node out-degree $\bar{k}_{out}$, (c) variance of node out-degree $\text{Var}(k_{out})$, and (d) the network diameter $\psi$ plotted against the length of the time series $N$. Each curve corresponds to ordinal networks constructed from one of five $x$-component time series generated by the Rössler system in different dynamical regimes with bifurcation parameters $\alpha = 0.37$ (period-2), $\alpha = 0.41$ (period-3), $\alpha = 0.39$ (multiband chaos), $\alpha = 0.4$ (broadband chaos) and $\alpha = 0.42$ (broadband chaos) respectively. The embedding dimension is set at $m = 8$ and the lag is set at $\tau = 9$. 

18
Quantifying changes in chaotic dynamics

The bifurcation diagram and largest Lyapunov exponent for the Rossler system are plotted against the bifurcation parameter in Figure 2.6a and 2.6b respectively. As the bifurcation parameter increases the system undergoes a period doubling route to chaos then passes through several periodic windows including the prominent period-3 window at $\alpha \approx 0.41$. The bifurcation diagram was generated using the maxima\(^1\) of the $x$-component time series which were enumerated using the function *extrema* from the TISEAN software package by Hegger et al. [1999]. This function performs quadratic interpolation on the data to obtain better estimates of the time series maxima. The largest Lyapunov exponent was also computed using TISEAN with algorithm by Kantz [1994] as implemented in the function *lyap.k*. This function returns the logarithm of the estimated exponential divergence of trajectories averaged over the reconstructed attractor with respect to the forward time evolution of the system. The gradient of the linear scaling region is the largest Lyapunov exponent $\lambda_1$ under the assumptions of determinism, ergodicity, stationary and, critically, the existence of said linear scaling region. The linear scaling region for the Rössler data was relatively consistent for the subset of $\alpha \in [0.37, 0.43]$ which generated chaotic time series. This enabled us to implement a script that automatically performed least squares regression analysis on the output from TISEAN for each of the 1201 time series and calculate $\lambda_1$ where a scaling region could be identified.

Figures 2.6d and 2.6e show that $\bar{k}_{out}$ and $\text{Var}(k_{out})$ both exhibit sensitivity to time series dynamics as they appear to track the relative change in $\lambda_1$ and detect all of the periodic windows. Similar but less effective tracking is apparent from $V$ in Figure 2.6c. This is to be expected, however, because $\lambda_1$, $\bar{k}_{out}$ and $\text{Var}(k_{out})$ all quantify complexity based on the evolution of trajectories. On the other hand, $V$ should be considered as a topological property as discussed in Section 2.2.1. The network diameter $\psi$ also exhibits sensitivity to dynamical changes as shown in Figure 2.6f. For example, a pronounced step change is evident in $\psi$ at the point of the first period doubling bifurcation and there are peaks at each of the periodic windows. While $\bar{k}_{out}$, $\text{Var}(k_{out})$ and $\psi$ all share the deficiency that they do not provide an absolute criteria for discriminating between periodic and chaotic dynamics, these results demonstrate that they have the potential to be useful as a simple tool for discriminating between time series observed from complex systems.

\(^1\)The set of maxima (or minima) of a time series is equivalent to a Poincaré section at the zero of its first derivative, as explained by Kantz and Schreiber [2004].
CHAPTER 2. UNWEIGHTED ORDINAL NETWORKS: SIMPLE STOCHASTIC MODELS FOR CAPTURING COMPLEX DYNAMICS

Figure 2.6: Basic measures computed on ordinal networks constructed from $x$-component time series generated by the Rössler system for bifurcation parameter $\alpha \in [0.37, 0.43]$. For reference we plot (a) the bifurcation diagram, and (b) the largest Lyapunov exponent $\lambda_1$. Network measures shown are (c) the number of nodes in the network $V$, (d) mean node out-degree $\bar{k}_{out}$, (e) variance of node out-degree $\text{Var}(k_{out})$, and (f) the network diameter $\psi$ plotted against the bifurcation parameter $\alpha$. The embedding dimension and lag parameters are set at $m = 8$ and $\tau = 9$ respectively.
A brief remark on the effect of noise

Consider the noise affected time series $y = \{y_n\}_{n=1}^N$ given by:

$$y = x + w,$$  \hspace{1cm} (2.5)

where $w = \{w_n\}_{n=1}^N$ is a series of independent and identically distributed observations drawn from the normal distribution $N(0, \sigma_{\text{noise}}^2)$. The noise level is expressed as a signal to noise ratio in decibels as

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left( \frac{\sigma_x^2}{\sigma_{\text{noise}}^2} \right),$$  \hspace{1cm} (2.6)

where $\sigma_x^2$ is the variance of the time series $x$. When the time series is corrupted with additive white Gaussian noise, $\bar{k}_{\text{out}}, \text{Var}(k_{\text{out}})$ and $\psi$ become far less effective for quantifying different time series dynamics. Measures of degree are global averages of local node properties which represent transitional possibilities for a single time step, hence these measures will be sensitive to noise. The network diameter will also be sensitive to noise depending on attractor topology. However, from a qualitative perspective, Figures 2.7 and 2.8 show that the distinct ring and band structures which we observed in the noiseless case for periodic and chaotic dynamics respectively are maintained even in the presence of noise.

**Figure 2.7:** Four ordinal networks constructed from the same periodic time series with increasing levels of additive white Gaussian noise. The underlying noiseless time series was generated by the Rössler system in a period-3 regime with bifurcation parameter $\alpha = 0.41$. The additive noise level expressed as $\text{SNR}_{\text{dB}}$ (Equation 2.6) is (a) 23.01 dB, (b) 20 dB, (c) 13.01 dB, and (d) 10 dB respectively. The embedding dimension and lag parameters are set at $m = 10$ and $\tau = 9$ respectively.

**Figure 2.8:** Four ordinal networks constructed from the same chaotic time series with increasing levels of additive white Gaussian noise. The underlying noiseless time series was generated by the Rössler system in fully developed chaos with bifurcation parameter $\alpha = 0.40$. The additive noise level expressed as $\text{SNR}_{\text{dB}}$ (Equation 2.6) is (a) 23.01 dB, (b) 20 dB, (c) 13.01 dB, and (d) 10 dB respectively. The embedding dimension and lag parameters are set at $m = 10$ and $\tau = 9$ respectively.
CHAPTER 2. UNWEIGHTED ORDINAL NETWORKS: SIMPLE STOCHASTIC MODELS FOR CAPTURING COMPLEX DYNAMICS

2.3 Application: Chaotic voltage signal time series measured from a diode resonator circuit

We now present results of the method as applied to experimental time series data from an externally driven diode resonator circuit (see Figure 2.9) as reported by Jüngling et al. [2008]. Each time series comprises 65536 observations of the voltage $U_R$. Time series were recorded for 1000 evenly spaced values of the control parameter: the amplitude of the driving sinusoidal voltage $U_0 \in [3, 5]$ volts. The system begins in period-3 oscillation (Figure 2.10a) and undergoes a period doubling bifurcation into period-6 (Figure 2.10b). This is followed by a period doubling cascade into multiband chaos (Figure 2.10c) and then an interior crisis at $U_0 \approx 4.05$ volts into chaos-chaos intermittency (Grebogi et al. [1986]) (Figure 2.10d). Networks for these different dynamical regimes are shown in Figure 2.11. Embedding lag is set to $\tau = 8$. The period-3 and period-6 time series map to ring structures (Figure 2.11a and 2.11b). In the period-6 case it was necessary to select a large value for the embedding dimension, $m = 16$, to unfold the ring structure (i.e. remove degeneracies from the ordinal map) because of the closeness of trajectories in phase space. The multiband chaotic time series exhibits a narrow flat band structure with a clear pinch point that corresponds to the folding region on the attractor (Figure 2.11c). Following the interior crisis the ordinal network reflects the complex self intersecting manifold of the attractor (Figure 2.11c). We postulate that the band-like community structures spiralling in and out of the network visualisation correspond to dense bundles of trajectories in phase space.

![Figure 2.9: The diode resonator circuit.](image)

The size of the networks grows with $m$ in the same manner as the Rössler system with the exception that the periodic networks begin to grow at a similar rate to the chaotic networks after $m = 15$ (Figure 2.12a). This anomaly is reflected in the plots for $\bar{k}_{out}$, Var($k_{out}$) shown in Figures 2.12b and 2.12c respectively, which both have peaks in the periodic regimes for $m > 15$ but otherwise exhibit similar trends to the results for the periodic Rössler time series. These later peaks are likely due to low levels of noise in the experimental time series. In discretely sampled periodic dynamics without noise, increasing $m$ will increase the likelihood that an embedded state vector $z_n$ will contain a pair of time series observations $x_k$ and $x_l$ that are close to being in phase and, hence, $|x_k - x_l| < \delta$ for some small $\delta$. If either $x_k$ or $x_l$ are perturbed by noise with amplitude $\epsilon > \delta$, then there is a significant possibility that the ordinal symbol for the noise affected data will be different from the symbol for the true trajectory. The onset of this effect will be sudden in periodic dynamics once $m$ is large enough that $\Delta t (m - 1) \tau$ is approximately equal to an integer multiple of the signal period. For example, the
2.3. APPLICATION: CHAOTIC VOLTAGE SIGNAL TIME SERIES MEASURED FROM A DIODE RESONATOR CIRCUIT

Figure 2.10: 3-dimensional time delay reconstructions of phase space from time series measured from the diode resonator circuit in different dynamical regimes with bifurcation parameters (a) $U_0 = 3$ volts (period-3), (b) $U_0 = 3.7$ volts (period-6), (c) $U_0 = 4$ volts (multiband chaos) and (d) $U_0 = 4.5$ volts (broadband chaos) respectively. The embedding lag is set at $\tau = 8$.

Figure 2.11: Ordinal networks corresponding to the reconstructed attractors in Figure 2.10. The networks were constructed with embedding lag $\tau = 8$ and embedding dimensions (a) $m = 10$, (b) $m = 16$, (c) $m = 14$, and (d) $m = 10$ respectively.

period of the period-3 time series corresponding to the results in Figure 2.12a is almost exactly 60 samples which explains why the network structure becomes highly sensitive to noise when $m = 16$, because $(m - 1)\tau = 15 \times 8 = 120$. The results for the normalised diameter $\psi/V$ shown in Figure 2.12d against $m$ also mirror those for the Rössler system: notably that values computed for periodic data are larger than those corresponding to chaos; and that chaos causes faster convergence to small values than periodic dynamics. When $m > 15$, results for $\psi/V$ computed for periodic time series decline rapidly and become indistinguishable from the results for chaotic time series. This implies a sudden increase in the complexity of the model and likely occurs due to low levels of noise in the data as discussed above.

The bifurcation diagram and largest Lyapunov exponent $\lambda_1$ of the data set are plotted in Figures 2.13a and 2.13b respectively. These were computed using the TISEAN package as per the methods described in Section 2.2.2. We generated networks for each time series with embedding dimension $m = 8$ and embedding lag $\tau = 8$. Network measures are plotted against the bifurcation parameter $U_0$ in Figures 2.13c-f. The size of the network exhibits sensitivity to both the period doubling bifurcation at $U_0 \approx 3.6$ volts, the period doubling cascade to chaos for approximately $U_0 \in [3.8, 4.05]$ volts, and undergoes a step change at the interior crisis, reflecting the filling of the attractor. As was the case for the Rössler system data, $k_{out}$, Var($k_{out}$) provide robust tracking of dynamical change similar to $\lambda_1$, and also appear sensitive to the period doubling bifurcation and the interior crisis. Network diameter $\psi$ undergoes a clearly discernible step change at the interior crisis. This result can be understood in terms of the relationship between the networks and phase space as follows. Additional nodes and edges
are created immediately after the crisis, corresponding to the intermittent chaotic trajectories that begin to fill the space between the bands of the pre-crisis attractor in phase space. These new nodes and edges become shortcuts in the network which significantly reduce $\psi$. The spike in $\psi$ at the step change corresponds to the small number of time series which have only a limited number of trajectories in between the bands of the pre-crisis attractor because they are in the immediate vicinity of the crisis. These trajectories will form new subgraph strands in the network which are only connected to the main structure where they leave and rejoin the bands of the pre-crisis attractor, and hence will significantly increase the network diameter. Moreover, these subgraph strands will have a far lower $\bar{k}_{out}$, $\text{Var}(k_{out})$ than the remainder of the network, hence why the value for mean out degree and degree variance also dips at the interior crisis.

Figure 2.12: (a) The number of nodes in the network $V$, (b) mean node out-degree $\bar{k}_{out}$, (c) variance of node out-degree $\text{Var}(k_{out})$, and (d) the network diameter that has been normalised with respect the size of the network $\psi/V$ plotted against the embedding dimension $m$. Each curve corresponds to ordinal networks constructed from one of time series measured from the diode resonator circuit in different dynamical regimes with bifurcation parameters $U_0 = 3$ volts (period-3), $U_0 = 3.7$ volts (period-6), $U_0 = 4$ volts (multiband chaos) and $U_0 = 4.5$ volts (broadband chaos) respectively. The embedding lag is set at $\tau = 8$. 

24
2.3. APPLICATION: CHAOTIC VOLTAGE SIGNAL TIME SERIES MEASURED FROM A DIODE RESONATOR CIRCUIT

Figure 2.13: Basic measures computed on ordinal networks constructed from time series measured from the diode resonator circuit with bifurcation parameter $U_0 \in [3, 5]$ volts. For reference we plot (a) the bifurcation diagram, and (b) the largest Lyapunov exponent $\lambda_1$. Network measures shown are (c) the number of nodes in the network $V$, (d) mean node out-degree $\bar{k}_{out}$, (e) variance of node out-degree $\text{Var}(k_{out})$, and (f) the network diameter $\psi$ plotted against the bifurcation parameter $U_0$. The embedding dimension and lag parameters are set at $m = 8$ and $\tau = 8$ respectively.
2.4 Chapter summary

In this chapter we defined and investigated properties of unweighted ordinal networks constructed from univariate time series data. The network topology of these simple stochastic models describes an approximation of the deterministic temporal relationships between symbolic states which are enumerated from data by the application of an ordinal partition to reconstructed phase space. Node out-degree and its mean value over the network can be interpreted as measures of local and global transitional complexity, respectively. For time series of some fixed length, the network size, variance of node out-degree and the network diameter can be used as relative measures which each quantify a different aspect of the global complexity of the data: network size is a coarse measure of the complexity of the empirical probability distribution of the symbolic dynamics; variance of node out-degree is related to the heterogeneity of the attractor; and diameter exhibits sensitivity to small changes in dynamics and intermittency.

Selecting the method’s two parameters, the embedding dimension and embedding lag, is a topic that will be addressed in detail in Chapter 3. However, in this chapter, our numerical investigation demonstrated that ordinal network analysis can be effective when using traditional criteria for selecting embedding lag such as using the first zero of the autocorrelation function. We specified the embedding dimension to maximise the amount of information encoded in the network based on peaks in the mean and variance of node out-degree, taking into account limitations imposed by finite data and sampling rate.

The central argument of this chapter has been that ordinal network structure is dependent on time series dynamics. To support this hypothesis we constructed ordinal networks and quantified network structure with simple measures for discrete sampled time series data from the low dimensional continuous chaotic Rössler system and experimental data from a diode resonator circuit. Our results demonstrated that network measures track the relative change in the largest Lyapunov exponent and detect periodic windows as these systems bifurcate over the domain of their respective control parameters. Of particular interest was our finding that network diameter was highly sensitive to the interior crisis at the boundary between multiband chaos to chaos-chaos intermittency in the diode resonator circuit data.

The primary disadvantage of using unweighted rather than weighted networks when constructing a Markov chain approximation of a dynamical system, is that the probability distributions of the symbolic dynamics may be severely distorted. This statement applies to both the probability of observing a given state, and the conditional probability of observing a specific pair or trajectory of symbolic states given an initial condition. In the chapters to follow we therefore continue our investigations using weighted ordinal networks. By drawing upon concepts from information theory and classical Markov chain theory, we propose refined measures for the express purpose of quantifying network topology for time series analysis.
Chapter 3

Weighted Ordinal Networks for Quantifying Transitional Complexity on Multiple Time Scales

ABSTRACT

Our aim in this chapter is to refine the ordinal network model by considering not only the existence of a transition between a pair of symbolic states, as in Chapter 2, but also incorporating the best estimate of the probability of each transition that can be extracted from the data. This can be achieved by weighting each network edge by the frequency with which each transition occurs in the ordinal symbolic dynamics. We proceed to examine the nature of an ordinal partition under the assumption of determinism with regard to reconstructed phase space, and discuss the implications for parameter selection and practical application of the method. The latter discussion includes commentary on the phenomenon that we call node aliasing, which can significantly affect measures of the connectivity or transitional structure on the networks, but can be minimised by suitably interpolating the data. Furthermore, we present an argument for ordinal networks as a generalised model for ordinal analysis and show that permutation entropy can be interpreted as the Shannon entropy of the stationary distribution of the Markov chain given by the network’s adjacency matrix. The central ideas we introduce in this chapter are two novel information theoretic measures of transitional complexity which we call local node out-link entropy and global node out-link entropy, and an accompanying scheme for multiscale time series analysis.

To validate our theoretical discussion we have undertaken a comparative test of permutation entropy, conditional entropy of ordinal patterns and global node out-link entropy using numerically generated time series from the Rössler system. We then apply our novel measure in a multiscale ordinal network analysis for discrimination between short time electrocardiogram recordings characterised by normal sinus rhythm, ventricular tachycardia and ventricular fibrillation respectively, and for the investigation of age-related effects in cardiac dynamics using interbeat interval time series from long time electrocardiograms.
3.1 Constructing weighted ordinal networks from univariate time series

To construct a weighted ordinal network from the uniformly sampled univariate time series \( x = \{x_n\}_{n=1}^N \) we first enumerate the ordinal symbolic dynamics \( s = \{s_n\}_{n=1}^{N-(m-1)\tau} \) for the symbols \( s_n \in \mathcal{S} \), where \( \mathcal{S} \) is the unique set of ordinal symbols (alternatively, the symbolic alphabet) as enumerated from the time series. This is done by precisely the same method defined for unweighted ordinal networks in Chapter 2 Section 2.1. In brief, this entails embedding the data in \( m \)-dimensions with lag \( \tau \) and then applying an ordinal partition to define a map between reconstructed phase space and the symbolic alphabet \( \mathcal{S} \). Typically we select \( m \) such that \( |\mathcal{S}| \ll N \) as discussed further in Section 3.2.1. Continuing as per the method for the unweighted case, we map the symbolic alphabet \( \mathcal{S} \) to the set of nodes \( \mathcal{V} \) in the network \( G = \{\mathcal{V}, E\} \) using the bijective operator \( \Gamma : \mathcal{S} \rightarrow \mathcal{V} \). Recall from Chapter 2 that the network \( G \) is represented by the adjacency matrix \( A \) with elements \( a_{i,j} \). As the nomenclature would suggest, the property which distinguishes weighted and unweighted ordinal networks is that the set of edges \( E \) are not binary connections but, rather, have weight given by the expression:

\[
a_{i,j} = |\{n \in \{1, 2, ..., (N - (m - 1)\tau)\} : (\Gamma(s_n), \Gamma(s_{n+1})) = (i, j)\}|, \text{ given } s_n \in \mathcal{S} \text{ and } i, j \in \mathcal{V}, \quad (3.1)
\]

where \( a_{i,j} \) is a non-negative integer. This means that a directed edge is placed from node \( i \) to node \( j \) with weight equal to the number of times this transition occurs in the symbol sequence \( s \). Note that \( a_{i,j} = 0 \) implies that there is no edge from node \( i \) to \( j \). A weighted ordinal network is a more complete Markov model of the time series than the unweighted networks which we investigated in Chapter 2 because the edge weights encode the additional information necessary to accurately compute estimates of the stationary and conditional probabilities of the symbolic dynamics, as will be discussed in Section 3.4. For this purpose, it is also critical that self edges are included in the definition of a weighted ordinal network.

3.2 Parameter selection

As is widely understood in the field of nonlinear time series analysis and recently summarised by Bradley and Kantz [2015], the selection of model parameters can have a significant effect on results and therefore must be treated with care. In their seminal paper defining ordinal partitions, Bandt and Pompe [2002] asserted that a partition should arise “naturally” from the data with no additional model assumptions. However, an ordinal partition is parameterised by an embedding dimension \( m \) and an embedding lag \( \tau \). Whilst it is advantageous that these parameters alone are sufficient to fully describe the partition, appropriate selection of \( m \) and \( \tau \) necessarily requires model assumptions. If it is assumed that some component of the time series dynamics is deterministic then a phase space interpretation of the ordinal partition, as first discussed by Groth [2005], is arguably the most valid approach. The following discussion addresses parameter selection for an ordinal partition in light of this interpretation.
3.2. PARAMETER SELECTION

Figure 3.1: An illustration of an ordinal partition applied to the $x$-component time series of the Rössler system which has been embedded with dimension $m = 3$. The set of coloured points on each plot of the reconstructed attractor fall within one of the 6 disjoint partition elements as given by the corresponding ordinal symbol shown below.

3.2.1 Embedding dimension

To apply an ordinal partition to a time series is to define a symbolic mapping where regions of the embedding phase space are uniquely mapped to the set of ordinal symbols. While these regions are strictly unbounded, they make up a set of disjoint subspaces which span the complete phase space. Groth [2005] identified that the boundaries between the regions are given by the $\binom{m}{2}$ inequalities between the dimensions of the embedding phase space that arise from the ordinal symbols. For example, the ordinal symbol $\{1, 2, 3\}$ corresponds to points embedded in $\mathbb{R}^3$ in the region with boundaries $x = y$, $x = z$ and $y = z$, and which satisfies $\{x > y > z\}$. Figure 3.1, shows an ordinal partition of $m = 3$ applied over the reconstructed phase space $a$ of a chaotic Rössler time series.

It should now be apparent that there exists a single fixed ordinal partition in $\mathbb{R}^m$ for each $m \in \mathbb{Z}^+$. Increasing $m$ serves to embed the time series $x$ over these fixed partitions in successively higher dimensions thereby unfolding the trajectories over an increasing number of possible ordinal subspaces. When embedding data for ordinal network analysis, it is important to select $m$ large enough such that different phase points are only mapped to the same symbol if they are genuine neighbours or else the
resulting network will contain degeneracies. Similar reasoning is presumably what led Amigó et al. [2014] to suggest the use of traditional methods from embedology to select \( m \) for ordinal analysis, such as computing false nearest neighbours. However, although there may be an analogy between unfolding over successively finer ordinal partitions and the computation of false nearest neighbours, the latter approach quantifies unfolding based on spherical \( \epsilon \)-balls in Euclidean space which are vastly different from the topology of the elements of an ordinal partition. This is not to say that traditional measures may not be useful for the purpose of selecting \( m \) for ordinal analysis, however, for ordinal networks we opt for a scheme more akin to the asymptotic invariant method as discussed in Small [2005], which is to compute some measure on the ordinal network and observe how it changes with the embedding dimension.

As will be shown in Section 3.5, for long and highly sampled time series from noiseless chaotic and periodic systems we observe that the information theoretic measure we call global node out-link entropy (defined in Section 3.4) will grow with \( m \) initially and then either decay to some constant value or continue growing to a constant or near constant value depending on the dynamics. The embedding dimension should be selected as the smallest \( m \) for which the change in global node out-link entropy with respect to \( m \) is small. This ensures that the majority of degeneracies have been resolved without choosing \( m \) so high as to encode redundant information. In the case of data which is less finely sampled or has some stochastic component, global node out-link entropy tends to grow to a maximum value and then decay. For this kind of data we select \( m \) based on the peak value of global node out-link entropy to maximise the amount of information encoded in the network.

In the wider literature surrounding ordinal analysis it is generally the case that smaller embedding dimensions are favoured and this is usually for one of two reasons. The first is that the \(|S|\) is bounded from above by \( m! \) and hence the symbolic alphabet can be very large. This can be problematic when computational resources are limited (although this upper bound is seldom approached for systems with a strong deterministic component, as will be discussed shortly). Secondly, to avoid undersampling when estimating the probability distribution of the symbolic dynamics, numerous studies cite the heuristic \( N \gg m! \) which originates from Bandt and Pompe [2002] and is supported by the subsequent work of Staniek and Lehnertz [2007] regarding parameter selection for permutation entropy measurements. An even more conservative constraint \( N \gg 5m! \) was proposed by Amigó et al. [2008]. Such criteria limit the range of the embedding dimension to \( 3 \leq m \leq 7 \) for most practical applications with experimental data. This condition is a necessary one when using ordinal analysis to test for evidence of determinism in noisy time series using forbidden patterns, a concept first proposed by Amigó et al. [2007], where it is critical that the system under observation is sufficiently sampled so only true forbidden patterns can be observed in the data (see also Amigó et al. [2008, 2010]; Rosso et al. [2012]). Forbidden patterns are the set of ordinal symbols that cannot occur in certain deterministic systems due to the fundamental properties of the dynamics. Amigó et al. [2010] analytically proved the existence of forbidden patterns in specific one dimensional maps and provided strong evidence for their existence in higher dimensional maps using numerical results. While false forbidden patterns will be observed if a system is undersampled, true forbidden patterns will dominate \(|S|\) as \( m \) increases for systems with a strong deterministic component such that \(|S|\) is generally \( \ll m! \). Therefore, when using ordinal analysis to measure the relative differences between two systems for which the assumption of a dominant
3.2. PARAMETER SELECTION

deterministic component in dynamics is reasonable, as is the case in this thesis, it is sufficient that \( N \) be large enough to sample the set of allowable patterns, not the entire set of possible permutations. While it is generally the case that we do not have knowledge of this set, it is fair to conclude that a time series is sufficiently long if the corresponding ordinal network converges on a stable structure for the truncated time series \( \{x_n\}_{n=1}^k \) with \( k \ll N \) with respect to increasing \( k \).

3.2.2 Embedding lag

In the theoretical limit of infinite data, the invariant properties of an embedded attractor should not depend on the choice of embedding lag, however Bradley and Kantz [2015] point out that this parameter can affect results when estimating invariants from finite experimental data. Since Cao et al. [2004] generalised the original definition of the ordinal symbolic mapping procedure to include embedding lag, numerous studies have investigated its effect in the context of ordinal analysis. Early work by Bandt and Shiha [2007] demonstrated that ordinal pattern statistics track the autocorrelation function of a sun spot time series. More rigorous research by Zunino et al. [2010] and Soriano et al. [2011] has shown that permutation entropy and permutation statistical complexity are both closely related to the intrinsic time scale of a system for the chaotic Mackey-Glass equations and experimental laser systems respectively. By using short ordinal symbols as bio-markers in the study of human cardiac dynamics, Parlitz et al. [2012] found that the symbols were most effective as statistical classifiers when the embedding lag was set to a characteristic frequency traditionally used in heart rate variability analysis. One finding of particular interest for researchers in experimental fields is the work of Zunino et al. [2012] where it was shown that, for data with both deterministic and stochastic components, selecting \( \tau \) on an appropriate time scale can maximise the contribution of determinism relative to the stochastic component in the resulting ordinal symbolic dynamics.

These observations are best explained in terms of the embedding phase space. Embedding lag \( \tau \) controls to what extent the dynamics are unfolded in \( \mathbb{R}^m \). The general consensus in embedology is that \( \tau \) should be selected such that the elements of the embedding vectors are as uncorrelated as possible to optimally unfold the attractor. In the context of ordinal analysis, an optimally unfolded attractor will occupy a maximal number of states over the partition (Groth [2005]; Riedl et al. [2013]) thereby maximising the useful dynamical information captured by the ordinal symbolic dynamics, as shown by Micco et al. [2012] who demonstrated that maximum values of permutation entropy statistical complexity corresponded to lags selected by traditional methods such as autocorrelation and mutual information.

The relationship between ordinal measures and \( \tau \) affords the possibility for multiscale analysis. This is achieved by computing ordinal measures from a single time series using several different lags to capture dynamics operating on different time scales. Such a scheme was applied to EEG signals by Keller et al. [2014] and for the analysis of both EEG and HRV data by Zunino et al. [2015]. We implement a similar multiscale approach in our investigation of ECG data in Section 3.6.

As an ancillary point, when selecting larger values of \( \tau \) for chaotic data it is important to consider that the elements of each vector will eventually become uncorrelated with increasing \( \tau \) and the symbolic dynamics will resemble a random variable (Popov et al. [2013]; Riedl et al. [2013]), although it is
plausible that the distribution of this random variable may still contain useful information about the
dynamics.

3.3 Practical considerations

3.3.1 Node aliasing

The phenomenon which we call node aliasing is a type of artefact that can occur in ordinal networks
when the sampling frequency of the time series is too low relative to the rate of evolution of the
trajectories and the length span of the elements of the ordinal partition in embedding phase space in
the direction tangential to the flow. When node aliasing occurs, it falsely implies local uncertainty in
the Markov chain which can give rise to misleading results.

For example, consider the trivial case of a continuous noiseless periodic time series oscillating at
frequency $\omega_0$ and sampled at $\omega \neq b\omega_0$, $\forall b \in \mathbb{N}$. It follows that there can exist an embedding vector $z_k$
which lies on an intermediate point of the trajectory between any two pairs of temporally adjacent
states $z_n$ and $z_{n+1}$. Since the edges of a ordinal network are allocated based on temporal succession,
depending on the particular choice of embedding parameters, the topology of the trajectories and the
sampling frequency of the data, it is possible that the connectivity patterns in the resulting network
may not be consistent with the true nature of the dynamics of the time series. One such scenario would
be that $z_n \rightarrow s_1$, $z_{n+1} \rightarrow s_3$, $z_{k-1} \rightarrow s_1$ and $z_k \rightarrow s_2$. The subgraph for nodes \{1, 2, 3\} is therefore:

\[
\begin{pmatrix}
1 & 2 & 3 \\
1 & 0 & 1 & 1 \\
2 & 0 & 0 & 0 \\
3 & 0 & 0 & 0
\end{pmatrix}
\]

We call this an instance of node aliasing because from node \{1\} it is impossible to distinguish between
the possible destination nodes \{2, 3\}. To correctly characterise the periodic dynamics, which contain
no such uncertainties, the correct subgraph should clearly be:

\[
\begin{pmatrix}
1 & 2 & 3 \\
1 & 0 & 1 & 0 \\
2 & 0 & 0 & 1 \\
3 & 0 & 0 & 0
\end{pmatrix}
\]

This can be obtained if our data is sufficiently over-sampled, but this creates two further problems.
The first is that in many applications we do not have the luxury of over-sampled data, however, we
show that simple interpolation of the data can suffice for the case of continuous chaotic time series in
Section 3.5. The second problem is that increasing the sampling frequency will increase the number of
self edges relative to the number of incoming and outgoing edges which distort certain local measures
of transitional complexity. We address this issue in Section 3.4.3 with our novel measure local node
out-link entropy.
3.4. QUANTIFYING THE COMPLEXITY OF ORDINAL NETWORKS

3.3.2 Data with coarse quantisation

Time series data maps to ordinal symbols based on the amplitude rank of the elements of embedded state vectors. A state vector with subset of the elements that are tied must be treated as a special case. The availability of high resolution analogue to digital converters in modern measurement devices means that ties are highly unlikely in most experimental time series data. In less common cases where ties are more frequent in data due to coarse quantisation, Bandt and Shiha [2007] and Zunino et al. [2010] suggest the addition of small amounts of white noise to break the equalities that will occur in the embedding vectors. This approach treats tied values inconsistently and the resulting symbolic dynamics will therefore contain an additional stochastic component. A more comprehensive method was proposed by Bian et al. [2012] whereby ties were permitted in the ordinal symbols. The authors tested this approach in an analysis of cardiac inter-beat interval time series and found that the entropy of the modified ordinal symbolic dynamics was effective for discriminating different age groups where traditional permutation entropy failed due to the frequent occurrence of tied values in the data. This scheme was also implemented by Makowiec et al. [2015]. The primary drawback with this approach is that symbols define subspaces with different dimensions. For example, the symbol \{1, 2, 3\} defines a three dimensional volume whereas \{1, 2, 2\} defines a two dimensional plane in the same three dimensional space. The resulting distribution of symbols is far more difficult to interpret as a measure of topology because the partition is highly non-uniform. It is for this reason that in this thesis we follow the standard convention for dealing with tied elements by assigning rank based on temporal order.

3.4 Quantifying the complexity of ordinal networks

3.4.1 Permutation entropy

Bandt and Pompe’s [2002] definition for the permutation entropy of a time series is the Shannon entropy of the corresponding set of ordinal symbols $s$:

$$h^{PE} = - \sum_i p_i \log p_i ,$$  \hspace{1cm} (3.2)

where the probability mass function $P(S = s_i) = p_i$ for $s_i \in S$ is estimated by counting the relative occurrence of each symbol in the symbolic dynamics $s$. Perhaps the most interesting property of permutation entropy from the theoretical perspective is that it is an upper bound for Kolmogorov-Sinai entropy for piecewise monotone interval maps and directly corresponds under certain conditions (see Amigó and Keller [2013]; Amigó [2012]; Keller and Sinn [2010]).

To demonstrate the connection between permutation entropy and ordinal networks, consider that the probability mass function is an estimate the stationary distribution of the Markov chain under the conditions that the network is irreducible and aperiodic. These conditions are discussed in detail in see Sections 4.1 and 4.2.3, but for now we shall take them as read. The stationary distribution can be
estimated by finding the stochastic matrix $P$ with elements
\[ p_{i,j} = \frac{a_{i,j}}{\sum_{k} a_{i,k}}, \tag{3.3} \]
where $a_{i,j}$ are the elements of the adjacency matrix $A$ of the ordinal network, and then calculating vector $q$ with elements $q_i$, which is the left eigenvector of $P$ corresponding to the eigenvalue $\lambda = 1$. By normalising we obtain the probability mass function
\[ p_i = \frac{q_i,\lambda=1}{\sum_k q_k,\lambda=1}. \tag{3.4} \]
Readers familiar with complex network theory will recognise this distribution can be interpreted as the left-eigenvector centrality of the stochastic matrix $P$ (Newman [2010]). It is also possible to estimate the stationary distribution from $A$ by taking
\[ p_i \approx \frac{\sum_j a_{i,j}}{\sum_k \sum_j a_{k,j}}, \tag{3.5} \]
where the sum of the edge weights for edges leaving node $i$ is divided by the sum of edge weights over the entire network. This is not strictly equivalent to counting the occurrences of ordinal symbols in $S$ due to the fact that Equation 3.5 counts occurrences based on the transition from one symbol to the next, and therefore does not count the final symbol in the sequence $s_{n-m+1}$.

### 3.4.2 Measures of transitional complexity

An inherent property of an ordinal network constructed from a discretely sampled continuous trajectory is that measures of the stationary distribution, including permutation entropy, will not necessarily distinguish periodic dynamics from chaotic dynamics or noise. A periodic limit cycle will often produce ordinal symbolic dynamics with a relatively uniform probability distribution over a set of symbolic states $S$ where $|S| > 1$. This results in a non-zero value of permutation entropy that could be misinterpreted as implying complexity despite the attractor being locally 1-dimensional everywhere. Permutation entropy will only be minimised for a system at a fixed point or for strictly monotonic signals. On the other hand, measures of transitional complexity quantify the local uncertainty of each state in the model and can therefore reveal additional information about the dynamics. For example, Froyland and Padberg-Gehle [2012] build Markov chains using a regular grid partition of phase space, and by taking the Shannon entropy for each box, compute what they call finite-time entropy which measures the nonlinear growth of a small set of initial conditions as the system evolves. This measure was shown to be useful for the detection of stable manifolds and local transport barriers, as well as being closely related to finite-time Lyapunov exponents. Furthermore, Ros et al. [2013] have demonstrated that node degree can be used as an index of predictability for discrete processes.

Recently Unakafov and Keller [2014] proposed a direct and logical extension to permutation entropy which they call the conditional entropy of ordinal patterns, henceforth referred to in this paper as conditional permutation entropy. They showed analytically and numerically that conditional
permutation entropy converges on Kolmogorov-Sinai entropy more quickly than permutation entropy. This measure was applied to electroencephalograms by Keller et al. [2014] for the classification of healthy subjects from those with epilepsy in both a normal state and during a seizure. Conditional permutation as defined by Unakafov and Keller [2014] is equivalent to

$$h_{CPE}^i = \sum_i \left( -p_i \sum_j p_{i,j} \log p_{i,j} \right), \quad (3.6)$$

where $p_i$ is computed as per Equation 3.2 and $p_{i,j}$ is the probability of a transition from $s_i$ to $s_j$ as estimated from $s$. Conditional permutation entropy is therefore the expected value of the entropy of node $i$ as averaged over the estimated stationary distribution of the ordinal network where we are referring to the concept of node entropy as defined by West et al. [2012], which is the inner summation in Equation 3.6. Within the context of ordinal network analysis, node entropy and its global average were first defined by Small [2013]. However, the variants of these measures used in that particular study assumed a uniform stationary distribution and were computed from unweighted ordinal networks built with non-overlapping embedding vectors. For completeness we note that conditional permutation entropy can also be described as the global average of finite-time entropy (Froyland and Padberg-Gehle [2012]) computed over an ordinal partition rather than a regular grid partition.

### 3.4.3 Local and global node out-link entropy of ordinal networks

Here we present the new measure central to this chapter. Consider a modified stochastic matrix $P^T$ of an ordinal network that excludes the possibility of self edges and hence has elements:

$$p^T_{i,j} = \begin{cases} 0 & \text{if } i = j, \\ \frac{a_{i,j}}{\sum_{k,k \neq i} a_{i,k}} & \text{if } i \neq j. \end{cases} \quad (3.7)$$

Taking the Shannon entropy of row $i$ gives the local node out-link entropy for node $i$:

$$h_{LNE}^i = - \sum_j p^T_{i,j} \log p^T_{i,j}. \quad (3.8)$$

It is possible that the final node in the symbolic dynamics $s$ has no outgoing edges in the ordinal network. If node $i$ has no out connections then we set $h_{LNE}^i = 0$ by definition to avoid singularities. The ordinal network mapping procedure guarantees that all other nodes will have nonzero out degree. By averaging over the network based on the stationary distribution as estimated by Equation 3.5 we obtain the expected value of the transitional complexity of $S$ which we call global node out-link entropy:

$$h_{GNE} = \sum_i p_i h_{LNE}^i. \quad (3.9)$$

The lower bound $h_{GNE} = 0$ implies that there is no uncertainty in the system. This corresponds to time series that are strictly monotonic or a symbolic dynamics $s$ that is strictly periodic. Note that periodic time series will only produce $h_{GNE} = 0$ if the ordinal map is free from degeneracies and aliasing effects. The maximum attainable entropy for a given node is bounded at $h_{LNE}^i \leq \log(k_i)$ where $k_i$ is the out
degree of node $i$. This corresponds to $p_{i,j} = 1/k_i$ and therefore maximum uncertainty. We choose not to normalise $h_i^{\text{LNE}}$ so as to preserve the absolute amount of information generated by each node before we take the expected value in Equation 3.9.

To derive an upper bound for $h_i^{\text{LNE}}$ which is independent of network topology, note that $V \leq m!$ which implies $k_i \leq (m! - 1)$ when self edges are removed. Maximum uncertainty for node $i$ arises when it is fully connected with $p_{i,j}^T = 1/(m! - 1)$ for all $j: j \neq i$, and therefore $h_i^{\text{LNE}} \leq \log(m! - 1)$. By extension, an ordinal network will have maximum uncertainty when it is fully connected with $V = m!$ and $p_i = 1/m!$ for all $i$. Assuming these conditions and substituting the upper bound for $h_i^{\text{LNE}}$ into Equation 3.9, the global node out-link entropy is also bounded at $h_G^{\text{GNE}} \leq \log(m! - 1)$. This corresponds to an infinitely long time series of uniform IID noise. Note that the upper bounds for permutation entropy and conditional permutation entropy can be derived via similar reasoning and are both equal to $\log(m!)$ (see Bandt and Pompe [2002]; Keller et al. [2014]). Upper bounds will be provided for reference in results from Sections 3.5 and 3.6.

The critical point of difference between our new measure $h_G^{\text{GNE}}$ and conditional permutation entropy is that we exclude self edges. In the case of a discrete sampled continuous dynamical system, the weights of the self edges encode the rate of evolution through each node in the ordinal network. Specifically, for a given node, dividing the weight of the self edge by the total weight of the outgoing edges excluding the self edge gives the average time that the system spends in that node each time it is visited, with respect to the sampling period of the data. It follows that, by increasing the sampling frequency, the weights of the self edges will increase. However, the weights of the outgoing edges should remain more or less the same if the base sampling frequency is high enough to minimise node aliasing. This relative change will skew the probability mass function for the transitional probabilities of a node in favour of the self edge and therefore cause conditional permutation entropy to converge to $h_C^{\text{CPE}} = 0$ as the sampling frequency tends to infinity. Using local and global node out-link entropy as computed from the modified stochastic matrix $P^T$ ensures that the resulting complexity measure is not dependent on the sampling frequency because we are separating information pertaining to transitional complexity from the rate of evolution and consider only the former.

### 3.5 Numerical investigation using the Rössler system

To demonstrate the effectiveness of global node out-link entropy $h_G^{\text{GNE}}$ in comparison with permutation entropy $h_P^{\text{PE}}$ and conditional permutation entropy $h_C^{\text{CPE}}$ we use numerically generated time series from the low dimensional chaotic system by Rössler [1976] (see Equation 2.2.2) with system parameters $\beta = 2, \gamma = 4$ and control parameter $\alpha \in [0.36, 0.43]$. Numerical solutions were computed using a fourth-fifth order Runge-Kutta algorithm with randomised initial conditions $\{x_0, y_0, z_0\} \in (0,1)$. We take a highly sampled time series of the $x$-component with transients removed and fixed data length sufficiently long such there are a total of 160 cycles when the system is in the period-2 regime. To find a suitable embedding lag we take the first zero of the autocorrelation function which is $\tau = 8$ when the data is down-sampled to a sampling period of $\Delta t = 0.2$ time units. In the following analysis we use this embedding lag and scale it directly with the sampling frequency as necessary to maintain an equivalent embedding.
Figure 3.2 shows the behaviour of the entropy measures with respect to the embedding dimension $m$ for a selection of periodic and chaotic time series. To show the effect of node aliasing, curves are plotted for both the original highly sampled data and down-sampled time series, with approximately 12000 and 60 points per cycle respectively. Note that the down-sampled data would be sufficient for most traditional methods of time series analysis, however, the reason we use over-sampled data is to investigate $h^{GNE}$ and $h^{CPE}$ in the case where node aliasing is minimal. Observe in Figure 3.2a that for highly sampled time series, $h^{GNE}$ converges to values close to zero for periodic systems and non-zero values for chaotic systems as expected. For $3 \leq m \leq 9$ the curves for $h^{GNE}$ reflect the unfolding of trajectories over the ordinal partition and the elimination of degeneracies with increasing $m$. All of the time series would appear to be suitably unfolded by $m = 10$ apart from the period-4 data. This particular time series was generated immediately following a period doubling bifurcation and contains trajectories that are very close in phase space. It is therefore necessary to select a larger $m$ to eliminate the degeneracies in the ordinal mapping. Henceforth, we set $m = 14$ for the subsequent analysis of the bifurcation spectrum of the Rössler system.

It can be seen in Figure 3.2b that for highly sampled data, $h^{CPE}$ is far less effective than $h^{GNE}$. While the $h^{CPE}$ curves for periodic and chaotic data do separate with increasing $m$, there is also a regular positive trend, unlike $h^{GNE}$ where we observe convergence. We have included results for $h^{PE}$ in Figure 3.2c for completeness. This plot demonstrates that $h^{PE}$ is not sensitive to oversampling and offers some degree of discrimination between periodic and chaotic dynamics. This distinction improves with $m$ as the ordinal partition becomes finer, but does not appear to converge until $m > 18$. It is unlikely that an embedding dimension this high would be suitable for experimental data due to noise and computational limitations.

Node aliasing effects are evident in the curves for the down-sampled time series with the periodic results converging to non-zero $h^{GNE}$ (Figure 3.2a). For chaotic time series, aliasing raises the maximum entropy for $3 \leq m \leq 10$, after which the curve trends down sharply and incorrectly implies less uncertainty for chaos than for periodic dynamics. The same trend can also be observed in the plot for $h^{CPE}$ from down-sampled data (Figure 3.2b). This is perhaps a less intuitive outcome of node aliasing and occurs because the effective size of each element in the ordinal partition reduces with increasing $m$ which amplifies the aliasing effect as the allocated network edges skip over more states. The resultant network encodes a dense bundle of trajectories as a single long isolated trajectory which does not accurately reflect the divergence of nearby initial conditions on the chaotic attractor.

Figure 3.3 shows the effect of oversampling on $h^{GNE}$ and $h^{CPE}$. In the case of $h^{GNE}$, increasing the sampling frequency reduces node aliasing, and values for periodic and chaotic dynamics converge to zero and fixed non-zero values respectively as the number of alias edges in the network becomes small (Figure 3.3a). The only exception is the period-4 time series which converges to a value close to zero for reasons as provided in the preceding discussion pertaining Figure 3.2a. The measured value of $h^{CPE}$ shown in Figure 3.3b, however, is strongly related to the sampling frequency and converges to zero for oversampled time series due to the dominance of self edges in the ordinal network as explained in Section 3.4. Figure 3.3 also demonstrates the important result that data sampled at a lower rate which is then interpolated with a simple cubic spline provides values of $h^{GNE}$ which are almost equal to those measured from the true highly sampled dynamics.
CHAPTER 3. WEIGHTED ORDINAL NETWORKS FOR QUANTIFYING TRANSITIONAL COMPLEXITY ON MULTIPLE TIME SCALES

Figure 3.2: (a) Global node entropy $h_{GNE}$, (b) conditional permutation entropy $h_{CPE}$ and (c) permutation entropy $h_{PE}$ plotted against the embedding dimension $m$ for a selection of time series generated by the Rössler system. The solid curves and black $y$-axis on the left of each panel show results for highly sampled time series with $\Delta t = 1 \times 10^{-3}$. The dashed transparent curves and grey $y$-axis on the right of each panel show results for the same set of time series after they have been down-sampled to $\Delta t = 2 \times 10^{-1}$. 
3.5. NUMERICAL INVESTIGATION USING THE RÖSSLER SYSTEM

Figure 3.3: (a) Global node entropy $h_{GNE}$ and (b) conditional permutation entropy $h_{CPE}$ plotted against the sampling period $\Delta t$ for a selection of time series generated by the Rössler system. The solid curves show results generated by taking highly sampled time series with $\Delta t = 1 \times 10^{-4}$ (the rightmost data point on each solid curve) and then downsampling. The dashed transparent curves show results generated by taking sparsely sampled time series with $\Delta t = 2 \times 10^{-1}$ (the leftmost data point on each curve) and interpolating the data with a cubic spline. Maximum possible entropy for both measures at $m = 14$ is approximately 36.343.
All three entropy measures under consideration have been computed from highly sampled time series for 1401 uniformly spaced values of \( \alpha \in [0.36, 0.43] \) and the results are plotted in Figure 3.4. A bifurcation plot and the largest Lyapunov exponent of the Rössler system with respect to the bifurcation parameter \( \alpha \) are included for reference. Our new measure \( h_{GNE} \) tracks the relative change in the largest Lyapunov exponent and detects periodic windows more effectively than both \( h_{CPE} \) and \( h_{PE} \). Periodic time series result in values of \( h_{GNE} \) close to theoretically assured baseline value of zero entropy, and chaotic time series give consistently higher values which would facilitate thresholding for discrimination of dynamics if desired. On the other hand, \( h_{CPE} \) and \( h_{PE} \) provide arbitrary and inconsistent results for complexity in periodic time series. For example, note the step change in \( h_{PE} \) after the first period doubling bifurcation. With regards to \( h_{CPE} \) specifically, the total range of \( h_{CPE} \) is small in comparison to the other measures which may imply that this measure could be less robust to degeneracies or other artefacts in the ordinal mapping process. Furthermore, \( h_{CPE} \) trends down in the chaotic regime for \( \alpha > 0.4 \) which is inconsistent with the dynamics where the attractor is becoming more chaotic and hence less predictable as quantified by the largest Lyapunov exponent. We acknowledge that for the particular embedding parameters used in this instance, \( h_{PE} \) follows the positive trend in the largest Lyapunov exponent through the chaotic regimes more effectively than \( h_{GNE} \), which remains relatively constant. However, it is important to recall that \( h_{GNE} \) and \( h_{PE} \) are measuring fundamentally different properties of the dynamics (see Section 3.4) and we therefore suggest computing both to ascertain which is more useful for a specific application. Finally, we report that the curve for \( h_{GNE} \) against \( \alpha \) is virtually identical for highly sampled data and for down-sampled data that has been re-interpolated back to the original sampling frequency (not shown) — a result that gives merit to our new measure despite the impracticality of obtaining significantly over-sampled data from experimental systems, because interpolation can be sufficient for minimising node aliasing while also preserving the dynamics.

3.6 Application: Electrocardiogram analysis

3.6.1 Classifying cardiac dynamics using ordinal network entropy

Description of data and methodology

In this section we apply ordinal network analysis, specifically the computation of global node out-link entropy, to a dataset of human electrocardiograms (ECGs) recorded at the Royal Infirmary of Edinburgh’s coronary care unit. Henceforth we refer to this as the RIE:CCU dataset. The dataset comprises 81 ECGs each 10000 points in length that have been sampled at 500 Hz with 10 bits resolution. The records come from 13 different patients and were originally measured to observe the evolution in cardiac dynamics from normal sinus rhythm (NSR) through to ventricular tachycardia (VT), and finally to ventricular fibrillation (VF). There are 31 records of NSR, 30 records of VT and 20 records of VF. For a thorough description of the data see Small et al. [2002]. In this study we use this dataset as a simple test of the ordinal network method for the classification of different dynamical behaviour. We do not apply any preprocessing to the data. Given the vast differences in the dynamics of ECG signals characterised by NSR, VT and VF, a method such as ours should provide statistically significant discrimination of the three pathological groups. The global node out-link entropy \( h_{GNE} \)
3.6. APPLICATION: ELECTROCARDIOGRAM ANALYSIS

Figure 3.4: (a) The largest Lyapunov exponent $\lambda_1$ of the Rössler system plotted against the bifurcation parameter $\alpha$ and overlaid on the bifurcation diagram as computed from the $x$-component time series using the $\text{lyap.k}$ and $\text{extrema}$ functions, respectively, from the TISEAN software package by Hegger et al. [1999]; and (b) global node out-link entropy $h^{\text{GNE}}$, (c) conditional permutation entropy $h^{\text{CPE}}$ and (d) permutation entropy $h^{\text{PE}}$ computed using the $x$-component time series and plotted against the bifurcation parameter $\alpha$. These results are for highly sampled time series with $\Delta t = 1 \times 10^{-3}$. Maximum possible entropy for all measures at $m = 14$ is approximately 36.343.
of each time series is computed for both a short and a long embedding lag, and the resulting two
dimensional vector constitutes a measure of multiscale complexity. Canonical correlation analysis is
then performed to quantify the level of discrimination between the pathological groups.

To select the embedding lag $\tau$ we begin with the very simple assumption that the mean resting
heart rate is 80 beats per minute, which equates to 375 samples per cycle. The short embedding lag
is set at $\tau = 20$ which is, very approximately speaking, on the order of a quarter period of any of
the individual PQRST components of the complete cycle. Hence, we have selected $\tau$ to unfold the
individual wave components over the ordinal partition in embedding phase space, or in other words,
each symbol captures the detailed shape of ECG wave in segments of approximately 0.2 to 0.6 cycles
for $5 \leq m \leq 10$. The long embedding lag is chosen to be one order of magnitude larger at $\tau = 200$
which will capture dynamics over segments of approximately 2 to 6 cycles for $5 \leq m \leq 10$ and therefore
encode information pertaining to the inter-cycle variability of the ECG signal in the ordinal network.

Figure 3.5 shows $h_{GNE}$ computed for $3 \leq m \leq 10$. For $\tau = 20$ (Figure 3.5a) we select $m = 8$
as an intermediate embedding dimension between the apparent maxima for NSR and VF records
at $m \approx 6$ and the maxima for VT records at $m \approx 9$. To maintain consistency in $h_{GNE}$ between
the short and long time scales we opt to also set $m = 8$ for $\tau = 200$. It is clear from Figure 3.5b
that this embedding dimension is larger than would be chosen based on the criteria of maximising
$h_{GNE}$. However, discrimination between the pathological groups is still evident for this combination of
embedding parameters.

![Figure 3.5](image_url)

**Figure 3.5:** Global node entropy $h_{GNE}$ for the RIE:CCU ECG dataset plotted against the embedding
dimension $m$ for (a) embedding lag $\tau = 20$, and (b) $\tau = 200$. Each datum is the result for a single
time series at each value of $m$. Results are labelled and grouped according to pathology: normal sinus
rhythm (NSR), ventricular tachycardia (VT), and ventricular fibrillation (VF).

**Results and discussion**

The NSR, VT and VF waveforms each have distinct characteristic relationships between $h_{GNE}$ and $\tau$
as shown in Figure 3.6. For the NSR records, $h_{GNE}$ gradually declines over $0 \leq \tau \leq 200$ which suggests
that these ECG signals have some degree of correlation between temporally adjacent cycles. While
some of the VT records exhibits this same gradual decline in the $h^{GNE}$ for $10 \leq \tau \leq 200$, the bulk of the data is characterised by an sudden drop in entropy over $0 \leq \tau \leq 10$ followed by a value with a relatively stable mean but a significant variance. The curves for VF records are similar except they decay over $0 \leq \tau \leq 50$ and the steady state value is lower and has far less variance with respect to $\tau$. The implication in both of these cases is that when each ordinal symbol is spanning the equivalent of as little as 0.3 and 1.2 cycles of normal sinus rhythm, for VT and VF data respectively, the elements of each embedding vector have become uncorrelated and the ordinal symbols behave like random variables. This result is intuitive as one would expect that these signals do not contain temporal correlation on long time scales because the dynamics are symptomatic of system failure. Figure 3.6 also provides further justification for our choice of $\tau$. The short embedding lag, $\tau = 20$, allows us to compare the data on a time scale where temporal correlation may be present, while the long lag, $\tau = 200$, provides a time scale where the symbolic dynamics for all of the data are behaving as random variables, and therefore we are comparing the distributions of these variables.

The two dimensional multiscale plot of $h^{GNE}$ and the corresponding box plot are provided in Figure 3.7 showing clear three way discrimination between the pathological groups. The results for the canonical correlation analysis given in Table 3.1 demonstrate quantitatively that there is high correlation between the pathological groups and multiscale $h^{GNE}$, and that there is strong statistically significant separation of the means of each group in two dimensions. While these results are all as expected in terms of relative classification, the actual values of $h^{GNE}$ present in each group raise questions about what aspect of complexity is being captured from the dynamics. For example, it may appear counter intuitive that an ECG record for VF is apparently less complex than for normal sinus rhythm. It is necessary to recognise that what is being measured here are the first-order state transition probabilities encoded in the ordinal network model. Ventricular fibrillation is characterised by extremely unpredictable time series that possibly result from high dimensional chaos or a stochastic process. If such highly irregular data is embedded (if it is even appropriate to do so given its potentially stochastic nature) and $m$ is selected such the time series is not long enough to ensure good sampling over the number of possible ordinal symbols $m!$, then the symbolic dynamics will likely comprise a series of long non-repeating trajectories with a low recurrence rate. The corresponding ordinal network will then comprise long chains of nodes with $k_{out} = 1 \implies h^{LNE} = 0$ that are connected by small number of hub nodes with $k_{out} > 1$ which arise from a state recurrences (see example in Figure 3.8). This degree of under-sampling can therefore result in lower values of $h^{GNE}$ than would be expected. While $h^{GNE}$ was effective as a statistic for relative classification in this particular analysis, it is not reasonable to infer that the complexity of the ordinal network reflects the complexity of the data. On the other hand, the prospect of inferring characteristics of a real system becomes more reasonable when long time series from controlled experimental conditions are available, as is the case for the data investigated in Section 3.6.2.

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>0.8061</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees of freedom for $\chi^2$ test statistic</td>
<td>2</td>
</tr>
<tr>
<td>Right-tail significance level for $\chi^2$ test statistic</td>
<td>$1.6986 \times 10^{-18}$</td>
</tr>
</tbody>
</table>
CHAPTER 3. WEIGHTED ORDINAL NETWORKS FOR QUANTIFYING TRANSITIONAL COMPLEXITY ON MULTIPLE TIME SCALES

Figure 3.6: Global node entropy $h_{GNE}$ for the RIE:CCU ECG dataset plotted against the embedding lag $\tau$. Each curve corresponds to a single time series. Results are grouped according to pathology. Maximum possible $h_{GNE}$ at $m = 8$ is approximately 15.299.
3.6. APPLICATION: ELECTROCARDIOGRAM ANALYSIS

Figure 3.7: (a) Multiscale scatter plot of global node out-link entropy $h^{GNE}$ for the RIE:CCU ECG dataset. The $x$ and $y$ axes correspond to $h^{GNE}$ computed with embedding lag $\tau = 20$ and $\tau = 200$ respectively. Each datum is the result for a single time series. Maximum possible $h^{GNE}$ at $m = 8$ is approximately 15.299. (b) Box plot for the scores from a canonical correlation analysis of the data computed using the `canoncorr` function in MATLAB R2014b.

Figure 3.8: Ordinal network constructed from an ECG time series record of ventricular fibrillation (VF). Embedding parameters are $m = 8$ and $\tau = 200$. 

45
3.6.2 Heart rate variability analysis for different age groups using ordinal network entropy

Description of data and methodology

The second application of our ordinal network method which we include in this work is the investigation of age-related effects in interbeat interval dynamics from ECGs. We have used data from the Fantasia database which is publicly available on PhysioNet (see Goldberger et al. [2000]) and was originally recorded for, and studied by Iyengar et al. [1996]. The authors of this paper attempted to produce high quality ECG recordings and isolate age as the only experimental variable to the extent that it was practical to do so, such that they could investigate differences in fractal scaling between age groups via detrend fluctuation analysis. The data comprises long time ECG recordings taken under controlled conditions from 20 healthy young subjects aged 21 to 43, and 20 elderly subjects aged 68 to 85. The elderly subjects were rigorously screened for medical conditions and subsequently also deemed to be healthy. There are equal numbers of females and males in each group. The records range in length from 1558559 to 2342528 observations and are sampled at 250Hz. The interbeat interval time series were computed from the ECG using automated beat annotations that had been confirmed by visual inspection. The reader should refer to Iyengar et al. [1996] for a detailed description of the data.

Interbeat interval data poses two problems for ordinal analysis. The first is that equal values in the embedding vectors are encountered often due low sampling frequencies. The second problem is that an interbeat interval time series can be considered as a sparsely and unevenly sampled record of the instantaneous period of the signal which will cause unpredictable node aliasing and degrade the network model of the dynamics. We address these issues by taking the original time series \( \{y(n)\}_{n=1}^{N} \) with elements that are the time indices of each R wave in sequence, and transforming it into the new time series \( \{\hat{y}(\hat{n})\}_{\hat{n}=1}^{\hat{y}(N)} \), with elements

\[
\hat{y}(y(n)) = \frac{1}{y(n+1) - y(n)}, \text{ for } n \in [1, N], \tag{3.10}
\]

To estimate \( \hat{y}(\hat{n}) \) for all other values of \( \hat{n} \in \{y(1), (y(1) + 1), (y(1) + 2), ..., y(N)\} \) that are not defined in Equation 3.10 we use cubic spline interpolation so that the resulting sample rate is the same as the original ECG at 250Hz. The transformed time series \( \hat{y}(\hat{n}) \) serves as an approximation of the continuous heart rate signal with uniformly spaced observations at a high sampling frequency. Since it makes sense to consider embedding lag in terms of number of heart beats, as given by the observed interbeat interval data, we define a new lag variable

\[
\hat{\tau} = \frac{1}{N-1} \sum_{n=1}^{N-1} y(n+1) - y(n), \tag{3.11}
\]

hence, our fundamental embedding lag is the mean period of one heart beat.

For multiscale analysis with global node out-link entropy \( h_{GNE} \) we use \( \hat{\tau} = 1 \) as the short embedding lag to capture information about sequences of temporally adjacent beats. The long time scale embedding lag is set arbitrarily at \( \hat{\tau} = 10 \) to be one order of magnitude larger. Figure 3.9 shows that the growth and decay of global node out-link entropy \( h_{GNE} \) is consistent between the age groups and has a clear
3.6. APPLICATION: ELECTROCARDIOGRAM ANALYSIS

Figure 3.9: Global node out-link entropy $h^{GNE}$ for the interbeat interval data from the PhysioNet Fantasia ECG dataset plotted against the embedding dimension $m$ for equivalent embedding lag (a) $\hat{\tau} = 1$, and (b) $\hat{\tau} = 10$ (see Equation 3.10 for the definition of $\hat{\tau}$). Each datum is the result for a single time series at each value of $m$.

maximum for both short and long $\hat{\tau}$ respectively. It is therefore a simple task to select the embedding dimension corresponding to the peak values of $h^{GNE}$ to maximise the transitional information encoded in the ordinal network, hence we have $m = 7$ for the $\hat{\tau} = 1$ and $m = 6$ for the $\hat{\tau} = 10$.

Results and discussion

It has been shown by Bian et al. [2012] that standard permutation entropy is unable to discriminate between the age groups based on the interbeat interval time series from the Fantasia database. We report the same result and, furthermore, that conditional permutation entropy does not effectively discriminate based on the interbeat interval time series nor with the approximation of the continuous heart rate (figures omitted for brevity). On the other hand, Figure 3.10a shows distinct clustering of the two age groups using $h^{GNE}$. Despite small overlap between the clusters, results from a one way ANOVA test and canonical correlation analysis given in Table 3.2 demonstrate that $h^{GNE}$ provides statistically significant discrimination of the means of each group with $p$-value $< 5 \times 10^{-04}$.

In the original investigation of the Fantasia interbeat interval data, Iyengar et al. [1996] found characteristic differences in the fractal scaling exponents interbeat interval time series based on subject age. It was shown that the fractal scaling exponent for records from elderly subjects comprised two distinct scaling regions with a bisection at approximately 40 beats. A two dimensional plot of the short and long range correlations as quantified by the fractal scaling exponent demonstrated that records from elderly subjects generally had higher short range correlation and lower long range correlation. The authors also reported that the variability in the fractal scaling exponent was significantly higher in the elderly subject group. Our choice of parameters results in ordinal symbols which span approximately 7 beats and 60 beats for short and long time scales respectively, hence straddling the formerly reported division of time scales. In turn, Figure 3.10a shows that the transitional complexity of the ordinal networks, as quantified by our measure $h^{GNE}$, is generally higher for elderly subjects on short time
Figure 3.10: (a) Multiscale scatter plot of global node out-link entropy $h^{GNE}$ for the interbeat interval data from the PhysioNet Fantasia ECG dataset. The $x$ and $y$ axes correspond to $h^{GNE}$ computed with equivalent embedding lag $\hat{\tau} = 1$ and $\hat{\tau} = 10$ respectively (see Equation 3.10 for the definition of $\hat{\tau}$). Each datum is the result for a single time series. Maximum possible $h^{GNE}$ is approximately 9.4898 and 12.299 for $m = 6$ and $m = 7$ respectively. (b) Box plot for a 1D projection of the multiscale scatter plot onto the vector between the means of each subject group.

scales, lower on long times scales, and has significantly greater variability than can be observed for young subjects, just as per the fractal scaling exponent as reported in the original study. While we refrain from any claim that $h^{GNE}$ is somehow equivalent to the fractal scaling exponent, it is by no means implausible that fractal correlation could manifest as complex transitional structures in ordinal networks. It is therefore reasonable to conclude that the results of our analysis support the findings from Iyengar et al. [1996].

Table 3.2: Statistical analysis of ordinal network results for the interbeat interval data from the Physionet Fantasia ECG dataset including: the significance level ($p$-value) of a one way ANOVA test on a 1D projection of the multiscale scatter plot (Figure 3.10a) onto the vector between the means of each subject group; and a canonical correlation analysis of the data computed using the `canoncorr` function in MATLAB R2014b.

<table>
<thead>
<tr>
<th>p-value for one way ANOVA</th>
<th>$2.0510 \times 10^{-04}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient</td>
<td>0.5810</td>
</tr>
<tr>
<td>Degrees of freedom for $\chi^2$ test statistic</td>
<td>2</td>
</tr>
<tr>
<td>Right-tail significance level for $\chi^2$ test statistic</td>
<td>$4.9061 \times 10^{-04}$</td>
</tr>
</tbody>
</table>

3.7 Chapter summary

In this chapter we introduced weighted ordinal networks as a more complete stochastic model of a deterministic time series than the unweighted models considered in Chapter 2. By defining edge weights as the frequency of transitions between states, it becomes possible to compute a better estimate of the probability distributions in the Markov chain which is being used to approximate the dynamics. Using
3.7. CHAPTER SUMMARY

This additional information that has been encoded in the network structure, we defined local and global node out-link entropy of ordinal networks as new measures to quantify the complexity of temporal structure in ordinal symbolic dynamics from time series. By assuming a deterministic component in the data and building on the existing literature, we have established a framework for conceptualising the ordinal mapping procedure in terms of reconstructed phase space. We also presented a case for weighted ordinal networks as a general model for ordinal analysis, specifically with respect to the widely used permutation entropy measure, which can be interpreted as the Shannon entropy of the stationary distribution of the Markov chain defined by the ordinal network under the assumptions of irreducibility and aperiodicity. Our comparative investigation using continuous chaotic time series from the Rössler system has demonstrated that global node out-link entropy tracks dynamical change through period doubling and periodic windows over a range of the bifurcation parameter. Qualitative assessment of the results confirm that our measure approaches a theoretical and intuitive lower bound of zero for periodic dynamics. This makes it arguably more useful than the existing and closely related measure of conditional permutation entropy which exhibits a sensitivity to the sampling frequency. Furthermore, our results indicate that global node out-link entropy is complimentary if not more effective than traditional permutation entropy for certain types of data, given that a fundamentally different property of the dynamics is being captured.

Considering previous studies relating to the relationships between embedding lag and ordinal analysis metrics, we proposed a simple scheme for multiscale analysis whereby we compute global node out-link entropy for different time scales by varying the lag parameter. In application, our method proved effective for discriminating between short time electrocardiogram recordings of normal sinus rhythm, ventricular tachycardia and ventricular fibrillation, and served to characterise differences in the transitional complexity of ordinal networks from interbeat interval time series on short and long time scales with respect to subject age. In the latter of these experimental investigations, we found that global node out-link entropy for healthy elderly subjects was higher on short time scales and lower for long time scales. This may suggest that normal ageing processes cause a change in the complexity of interbeat interval dynamics. We stress here the important distinction between measures of the model built from data and reality itself. Complexity in the ordinal network may not always be a true reflection of dynamical complexity in the underlying system. The amount of data available, quality of the data, and most importantly, the appropriateness of applying an ordinal partition to symbolise the data must all be taken into consideration when interpreting results from ordinal analysis. However, in this particular application the database under examination contained long time records taken under controlled conditions, the assumption of some deterministic process governing the underlying dynamics of the heart rate is reasonable, and our results appear to align the findings from the original study of the data. On the latter point, the seemingly analogous behaviour of global node out-link entropy and the fractal scaling exponent that we observed perhaps compels a more thorough investigation.

The primary limitations of ordinal network analysis are degeneracies in the mapping process and node aliasing. These effects are evident when comparing the curves for global node out-link entropy with the largest Lyapunov exponent, the latter of which still produces superior results for well behaved noiseless data. Degeneracies can be minimised to some extent by choosing sufficiently high embedding dimensions, however this will make the ordinal patterns less robust to noise, increase the detrimental

49
effect of node aliasing and can make computation impractical. We have shown that the node aliasing can be reduced by interpolation or by obtaining oversampled data. Using interpolated or oversampled data vastly improves results for global node entropy as a measure to discriminate between periodic and chaotic dynamics. However, in practice the availability of oversampled data depends on the nature of the system being measured and the measurement device being used. It may also be the case that the characteristic time scale of the system in question is not known a priori. Interpolation is also not a universally applicable solution because the interpolated data may differ from the true trajectory if the interpolation method (i.e. linear, cubic spline etc.) is not a good approximation of the time series dynamics.

Bradley and Kantz [2015] make it clear that due diligence should be observed when applying traditional methods of nonlinear time series analysis with respect to factors including model assumptions, parameter selection and numerical errors. The same is of course true for any form of ordinal analysis. Taking into account our current understanding of the theory and practical pitfalls, ordinal analysis can be applied as a form of data compression: filtering the data and retaining information of interest as described by Amigó et al. [2014]. We have shown that measures of ordinal network structure can be applied in this manner to classify biomedical time series data when used in conjunction with traditional statistical methods.
Chapter 4

Characterising time series complexity via inverse normalised ordinal network diameter

ABSTRACT

In this chapter we continue our investigation of measures on weighted ordinal networks to characterise the complexity of time series data from discretely sampled continuous systems. Having examined a local property of the network topology in the previous chapter, we now turn our attention to a global property, namely network diameter. In Section 4.1 we revisit the concept of ordinal networks as Markov models of time series, and establish our motivation for pursuing an investigation of network diameter within the context of Markov chain theory. Under conditions of irreducibility and aperiodicity, the long time probability of observing a particular state in a Markov chain from any initial state will converge on a unique stationary distribution. From existing literature we know that the time for the system to converge within some specified distance of the stationary distribution, called the mixing time, is bounded by the network diameter. We will derive a new diameter bound for the Markov mixing time based on the infimum of the stationary distribution.

We then discuss in Section 4.2 how diameter may provide an excessively conservative estimate of mixing time in ordinal networks which model time series from deterministic systems if there is a significant number of self edges. To improve the estimate, we define a modified ordinal network where edges are re-weighted equal to the average time spent in the previous node plus the time taken to transition to the next. In turn, we define a normalised measure of diameter on these modified networks to enable comparison between ordinal network models of any size, and postulate that the inverse of this measure should be related to the largest Lyapunov exponent for ordinal networks constructed from deterministic chaotic time series. We conclude this section with a discussion of the irreducibility of ordinal networks in theory and in practical applications under the assumption of ergodicity.

In Section 4.3 we perform a comparative numerical investigation of the inverse normalised network diameter and estimates of the largest Lyapunov exponent from time series data generated by the Rössler system, two variants of the Muthuswamy-Chua System, and the Chua circuit equations. Finally,
we apply the new measure in conjunction with a binary linear support vector machine with the aim of detecting and pre-detecting the onset of epileptic seizures from multi-variate electroencephalogram time series in Section 4.4.

4.1 Markov chain mixing time and network diameter

We begin this chapter by briefly reviewing several fundamentals of Markov chain theory as detailed in Levin et al. [2009] to define the mixing time of a Markov chain. Firstly, a Markov chain is called irreducible if the system state \( c \) can be reached from state \( b \) within a finite number of iterations of the transfer operator defined by the matrix of transitional probabilities \( P \) for any pair of states \( \{b, c\} \in S \) where \( S \) is the discrete state space of the system. In the context of ordinal networks, this is equivalent to stating that there exists a directed path of finite length between all possible node pair combinations \( \{i, j\} \in V \) from the network \( G = \{V, \mathcal{E}\} \) given by adjacency matrix \( A \), assuming that \( P \) is computed from \( A \) by Equation 3.3. Secondly, the period \( T(b) \) of a single state \( b \) in a Markov chain is defined as the greatest common denominator of the set of all possible return times (i.e. the g.c.d. of the set of all times \( n \in \mathbb{Z}^+ \) such that the probability of observing \( b \) after \( n \) steps of a random walk on the Markov chain which starts at state \( b \) is greater than zero). If \( T(b) > 1 \) for all \( b \in S \) then the Markov chain is called periodic, otherwise it is aperiodic.

Now consider state \( b \) at an initial time \( n = 0 \), expressed as a row vector of probabilities for the complete state space \( S \). To compute the probability of the system being in any given state \( c \in S \) after \( n \) steps we right multiply the row vector \( b \) by \( P^n \). One of the central results in Markov chain theory is that the product \( bP^n \) will converge to the unique stationary distribution \( \pi \) such that \( \pi = \pi P \) in the limit of \( n \to \infty \) if the Markov chain is both irreducible and aperiodic. The distance between \( bP^n \) and \( \pi \) is defined as:

\[
d(n) := \max_{b \in S} \|bP^n - \pi\|_{TV},
\]

where \( \|\nu - \nu\|_{TV} \) is the total variation distance between the probability distributions \( \nu \) and \( \nu \). The mixing time of a Markov chain can then be defined as:

\[
n_{\text{mix}}(\epsilon) := \min\{n : d(n) \leq \epsilon\},
\]

where \( \epsilon \) is generally taken as 1/4. Levin et al. [2009] show through a relatively simple proof that \( n_{\text{mix}}(\epsilon \leq 1/2) \leq \psi/2 \) where \( \psi \) is the diameter (equation 2.3) of the unweighted, undirected network given by adjacency matrix \( A \) with elements

\[
a_{i,j} = \begin{cases} 
1 & p_{i,j} + p_{j,i} > 0, \\
0 & \text{otherwise},
\end{cases}
\]

where \( p_{i,j} \) is the probability of a transition from state \( i \) to \( j \) given by \( P \) for all possible state pairs \( \{i, j\} \in S \).

The existence of a diameter bound on the Markov mixing time serves as the primary motivation for the work to follow in the remainder of this chapter. We propose that this relationship can also be
understood in terms of the equality:

$$d(\psi - 1) \geq \min\{\pi_c : c \in S\} > 0,$$  \hspace{1cm} (4.4)

where \(\pi_c\) is the probability of observing state \(c\) in \(\pi\) (i.e. as \(n \to \infty\)) and \(\psi\) is measured on the network as defined in Equation 4.3. This means that for \(n < \psi\) there always exists a non-zero distance \(\epsilon \geq \min\{\pi_c : c \in S\}\) between the distribution \(bP^n\) and \(\pi\) for any initial state \(b \in S\) which, therefore, bounds the mixing time by

$$n_{\text{mix}}(\min\{\pi_c : c \in S\}) \geq \psi.$$ \hspace{1cm} (4.5)

We prove this as follows. Consider the irreducible and aperiodic Markov chain defined by \(P\) with state space \(S\) and stationary distribution \(\pi\). Irreducibility implies network diameter \(\psi \leq |S|\). Therefore, a random walk on \(S\) governed by \(P\) of length \(n \leq \psi - 1\) for any initial state \(b \in S\) will span a maximum of \(|S| - 1\) states and, hence, the distribution \(bP^n\) will always contain at least one element with zero probability. This implies that

$$\|bP^n - \pi\|_{TV} \geq \min\{\pi_c : c \in S\},$$ \hspace{1cm} (4.6)

and, furthermore, \(\pi_c > 0\) for all \(c \in S\) because the Markov chain is irreducible. Substituting Equation 4.6 into Equation 4.1 leads to Equation 4.4 and, in turn, the bound on \(n_{\text{mix}}\) given in Equation 4.5.

Ros et al. [2013] proposed a related idea using unweighted, directed transition networks constructed from coarse grained continuous state space of chaotic maps. They define a time horizon for predictability \(n_{\text{max}} = \min\{n : a_{i,j}(n) = 1\}\) where \(a_{i,j}(n)\) are the elements of \(\Theta(A^n)\), \(A\) is the network adjacency matrix, and \(\Theta\) denotes the Heaviside step function. The authors make note of the implication that the transition matrix \(A^{n_{\text{max}}}\) (a full matrix of ones) provides no information about the future state of the system because all outcomes are equally likely. Their measure \(n_{\text{max}}\) is closely related to the network diameter \(\psi\). In fact, if one modifies the definition of diameter given in Equation 2.3 to include the distance from a node to itself (the shortest loop distance for a given node) then \(n_{\text{max}} = \psi\) for networks with no self edges.

### 4.2 Ordinal network based estimates of system mixing time

We have already briefly investigated the diameter of unweighted ordinal networks as a measure of complexity in Chapter 2. Our primary aim in this chapter is to investigate diameter based measures of mixing time when we incorporate the additional information contained in the network edge weights.

#### 4.2.1 Computing ordinal network diameter, the assumption of determinism, and the problem of self edges

The weighted network diameter is defined as:

$$\chi = \sup\{\bar{g}_{i,j}\} \text{ for all } \{i, j : i \neq j\} \in \mathcal{V},$$ \hspace{1cm} (4.7)
where \( g_{i,j} \) is the weighted geodesic or minimum length path between node \( i \) and node \( j \), and the weight of an edge is considered to be its length. Henceforth in this chapter we consider the weighted ordinal network adjacency matrix \( A \) with elements \( a_{i,j} \) as defined in Section 3.1. Note that computing the weighted diameter of \( A \) directly would produce an arithmetic sum of probabilities for a series of independent events and would therefore bear no relevance to the dynamical evolution of the system. Rather, we are interested in a measure with units of time that can be used to estimate mixing time. In the Markov chains for which we derived bounds in Section 4.1, an edge was assumed to have length equal to one unit of discrete time, and in this case weighted diameter \( \chi \) and unweighted diameter \( \psi \) are equivalent. However, this approach may discard significant information in ordinal networks or other Markov chains constructed from time series.

Consider the simple example where some deterministic time series \( x = \{x_n\}_{n=1}^N \) is mapped to the ordinal symbolic dynamics \( s = \{s_{n}\}_{n=1}^{N-(m-1)\tau+1} \) for \( s_n \in \mathcal{S} \). Assume that the nature of the deterministic dynamics is such that a specific symbol state \( b \in \mathcal{S} \) always appears twice in immediate succession in \( s \) before transitioning to the next symbol. This implies that the node in the ordinal network corresponding to \( b \) has a self edge with weight equal to some integer multiple of 2. However, by definition, a geodesic can never contain a self edge. Therefore, if the largest geodesic contains the node corresponding to \( b \) then the unweighted network diameter \( \psi \) (or weighted diameter \( \chi \) where \( a_{i,j} \in \{0,1\} \) for all \( i,j \in \mathcal{V} \)) will be an excessively conservative bound for the mixing time because it ignores the self edge on this node. This is simply a reflection of the fact that ordinal network models are stochastic and so can only approximate, not enforce, the determinism in the time series data from which they are constructed.

### 4.2.2 Improving estimates by re-weighting the ordinal network edges

In this section we propose a simple modification to the ordinal network adjacency matrix to incorporate the additional information contained in the distribution of self loops and obtain better estimates of mixing time. Recall from Section 3.4.3 that we can compute the mean time that the system spends in a given state \( b \in \mathcal{S} \) corresponding to node \( i \in \mathcal{V} \) each time that state is visited by dividing the weight of the self edge \( a_{i,i} \) by the sum of the weights of all other out-connected edges \( \{a_{i,k} : k \neq i, k \in \mathcal{V}\} \). We call this the mean node time for node \( i \). Therefore, we compute the modified adjacency matrix \( A^* \) with elements:

\[
a^*_{i,j} = \begin{cases} 
\frac{a_{i,i}}{\sum_{k,k \neq i} a_{i,k}} + 1 & a_{i,j} > 0, \\
0 & a_{i,j} = 0,
\end{cases}
\]

for all \( i,j \). (4.8)

This means that for all nodes \( i \in \mathcal{V} \), the weight of all of the out-connected edges for a given node \( i \) have weight equal to the mean node time plus a single time step to account for the time taken to traverse the edge to the subsequent node \( j \). The weighted diameter \( \chi \) of \( A^* \) will be a less conservative estimate of the mixing time. However, it must be noted that \( \chi \) is no longer strictly a lower bound based on Equation 4.5 with respect to the ordinal symbolic dynamics \( s \) because we have weighted the edges by the mean time taken to traverse a node, not the minimum possible time.
To enable comparison of mixing times between different systems that may manifest ordinal network models which vary in size, we define the normalised weighted ordinal network diameter as:

$$\hat{\chi} = \frac{\chi}{\gamma},$$  

(4.9)

where $\gamma$ is the normalisation constant

$$\gamma = \sum \left( \frac{a_{i,i}}{\sum_{k,k \neq i} a_{i,k}} + 1 \right) - \min \left( \frac{a_{i,i}}{\sum_{k,k \neq i} a_{i,k}} + 1 \right).$$  

(4.10)

The constant $\gamma$ can be interpreted as the longest possible geodesic path for a set of nodes $V$ (and the corresponding set of mean node times) for any possible edge configuration given by $E$. Intuitively, this geodesic path is a non-repeating sequence comprising all nodes $i \in V$ where the node with the shortest mean node time is located at the end of the sequence, hence $\gamma$ is an upper bound for $\chi$. Furthermore, given that both $\chi$ and $\gamma$ have units of time, $\hat{\chi}$ is therefore dimensionless.

For the remainder of this chapter we shall investigate the inverse normalised ordinal network diameter $\hat{\chi}^{-1}$ which is a dimensionless measure of mixing rate. The measure is bounded from below by $\hat{\chi}^{-1} \geq 1$ when $\chi = \gamma$ which implies either the slowest possible rate of mixing if the network is irreducible and aperiodic (i.e. the network has a single node $i$ with out-degree $k_{out}(i) = 2$, and all other nodes have $k_{out} = 1$), or that the network is irreducible but periodic and therefore does not mix. Fastest possible mixing occurs in the case where the network is fully connected and has no self edges. If we assume knowledge of the network size then this upper bound is $\hat{\chi}^{-1} \leq (V - 1)$ where $V = |V|$. Instead, if we are interested in a bound that is independent of the dynamics and only depends on the parameters then $\hat{\chi}^{-1} \leq (m! - 1)$.

### 4.2.3 Obtaining irreducible ordinal networks in practice

As discussed in Section 4.1, Markov chain mixing time can only be defined if the Markov chain is irreducible and aperiodic. We will now specify a single condition that can be easily tested which guarantees that an ordinal network is irreducible.

Consider the time series $x = \{x_n\}_{n=1}^{N}$ which is mapped to the ordinal symbol dynamics $s = \{s_n\}_{n=1}^{N-(m-1)r}$ for $s_n \in S$. A weighted or unweighted ordinal network $G = \{V, E\}$ given by adjacency matrix $A$ that has been constructed from $x$ is irreducible if and only if there exists $s_n = s_1$ and $s_k = s_N$ such that $n \geq k$ for $n, k \in 1, \ldots, N$. Equivalently, let $s_1 \xrightarrow{\Gamma} u$ and $s_N \xrightarrow{\Gamma} v$ for $u, v \in V$ where $\Gamma : S \rightarrow V$, then $G$ is irreducible if and only if there exists a finite length directed path from node $v$ to node $u$.

This statement can be proved by the following simple argument: by Equation 3.1, there always exists a directed edge with positive non-zero weight from the node corresponding to $s_n$ given by $\Gamma(s_n)$ and the node corresponding to $s_{n+1}$ given by $\Gamma(s_{n+1})$. By extension, there must exist a finite path from node $\Gamma(s_n)$ to the set of all nodes $\{\Gamma(s_k) : k > n\}$. Therefore, if there exists a finite length directed path from the node $v$ (corresponding to the final symbolic state $s_N$, given by $\Gamma(s_N)$), to node $u$ (corresponding to the initial symbolic state $s_1$, given by $\Gamma(s_1)$), then there must exist a finite path between all node pairs $\{i, j\} \in V$. Furthermore, if we assume that $x$ is ergodic then, by the Poincaré recurrence theorem (Marwan et al. [2007]), this path must exist for some sufficiently large but finite $N$. In this context, a
reasonable heuristic for sufficiently large \( N \) is \( N \gg m \) which is quoted numerous times in the ordinal analysis literature including Bandt and Pompe [2002], Staniek and Lehnertz [2007] and Amigó et al. [2014]. The inequality specifies that the number of observations in the time series must be far greater than the number of possible ordinal symbols, as discussed in Section 3.2.1. This condition will likely ensure that all possible ordinal symbols generated by a system will have occurred in the time series data sufficiently many times such that the empirical probability distribution of symbolic states is a good approximation of the true distribution.

In practice \( \mathbf{x} \) has some fixed finite length \( N \) and may be corrupted by noise or other transient perturbations which could result in an irreducible ordinal network. For example, consider that the dynamical system from which \( \mathbf{x} \) is observed is both ergodic and noiseless but was initialised in a state that did not lie on or close to the attracting manifold. While there likely exists some subset of the symbolic dynamics \( \{s_n\}_{n=k}^N \) for \( k > 1 \) which maps to an irreducible network for sufficiently large \( N \), the initial symbolic state \( s_1 \) may fall within an element of the ordinal partition that is never revisited in the dynamics. A similar problem can occur at the end of the time series if the system is noiseless but the dynamics are under-sampled (i.e. the data is too short) and the \( s_N \) corresponds to a trajectory that is visiting a new region of the attractor for the first time, or if noise has perturbed the trajectory off the attracting manifold. One can attempt to address these problems by truncating the symbolic dynamics from its beginning or end as required until the resulting ordinal network is irreducible. If this process discards too much data (or all of the data) then one can increase the likelihood of a state recurrence that will result in irreducibility by using a lower embedding dimension \( m \) and making the ordinal partition more coarse.

4.3 Numerical investigation

In this section we investigate the postulate that the inverse normalised ordinal network diameter \( \hat{\chi}^{-1} \) should be related to the largest Lyapunov exponent \( \lambda_1 \) for time series from deterministic systems characterised by chaotic mixing. Specifically, we undertake a comparative numerical investigation of \( \hat{\chi}^{-1} \) against estimates of the \( \lambda_1 \) using discretely sampled time series data generated by four different continuous chaotic systems: the Rössler system; the Muthuswamy-Chua system with quadratic nonlinearity (henceforth referred to as MC2); the Muthuswamy-Chua system with 4th-order polynomial nonlinearity (henceforth referred to as MC4); and the Chua circuit equations. For each system we generate an ensemble of time series for a range of the respective bifurcation parameter. The time series are finely sampled to minimise the effect of node aliasing on \( \hat{\chi}^{-1} \) (see Section 3.3.1). Sampling rates for each system are specified in their respective sub-sections to follow. Computation of the measure is repeated for down-sampled time series with various levels of additive white Gaussian noise (AWGN) expressed as SNR_{dB} as per Equations 2.5 and 2.6. For the following investigation we specify the variance of the time series \( \sigma_x^2 \) in Equation 2.6 as the maximum signal variance over the complete ensemble of time series for each system. This guarantees that the variance of the AWGN is constant over each ensemble and that the signal to noise ratio is less than or equal to the specified value.

Bifurcation diagrams are generated by finding the extrema of the \( x \)-component time series using the function \texttt{extrema} from the TISEAN software package by Hegger et al. [1999] (described in Section 2.2.2). The largest Lyapunov exponent \( \lambda_1 \) was estimated for the Rössler system from the \( x \)-component
time series using the TISEAN function \textit{lyap} \_\textit{k} which is an implementation of the algorithm by Kantz [1994] (described in Section 1.1). For the MC2, MC4 and Chua systems we used the TISEAN function \textit{lyap} \_\textit{spec} which is an implementation of the algorithm by Sano and Sawada [1985] and computes the spectrum of Lyapunov exponents using all time series components from the system or a delay coordinate embedding. Local linear models are used to reconstruct an approximation of the attractor’s tangent space from which the exponents can be computed directly. In this case we elect to use all three time series components generated by the respective systems of equations. The ordinal network diameter is computed using the algorithm by Dijkstra [1959] which is the fastest method of computing weighted network diameter in the general case with computation time \( O(E + V \log(V)) \) where \( E = |\mathcal{E}| \). Given sufficiently long and finely sampled data, as we are using here, we expect that larger \( m \) will result in the ordinal networks that are more accurate stochastic approximations of the time series dynamics. Network size \( V \) becomes the primary limiting factor in selecting \( m \) due to computation time. We therefore select \( m \) as large as possible such that \( V < 10000 \) for all (or at least most) of the time series from each respective system. Selected values of \( m \) are noted in the subsections for each respective system. Additional specific details about the systems, the time series and the network embedding parameters are also discussed within each subsection as follows.

4.3.1 The Rössler system

The Rössler system is defined in Equation 2.4. System parameters are \( \beta = 2, \gamma = 4 \), and bifurcation parameter \( \alpha \in [0.36, 0.42] \). The system is integrated from randomised initial conditions \( \{x_0, y_0, z_0\} \in (0, 1) \) using a fourth-fifth order Runge-Kutta method to generate finely sampled \( x \)-component time series with length \( N = 2000000 \) (after transients have been removed) at a sampling period of \( \Delta t = 0.001 \). This corresponds to approximately 6100 points per cycle. To obtain coarsely sampled data, the time series are down-sampled by a factor of 100.

The ordinal networks are constructed with embedding lag \( \tau = 1600 \) for finely sampled time series and, equivalently, \( \tau = 16 \) for coarsely sample time series. These lags correspond approximately to the first zero of the time series auto-correlation function which is also approximately one quarter of the signal period. The embedding dimension is set at \( m = 12 \) for noiseless data and \( m = 10 \) for data with AWGN.

An example of a chaotic Rössler attractor and its corresponding ordinal network are shown in Figure 4.1. Results for the computation of \( \lambda_1 \) and \( \hat{\chi}^{-1} \) are shown in Figure 4.2. The left and right \( y \)-axes of Figure 4.2b, corresponding to the curves for the two respective measures, are scaled to enable comparison between \( \lambda_1 \) which has units \([t^{-1}]\), and \( \hat{\chi}^{-1} \) which is dimensionless. The theoretical lower bounds \( \lambda_1 \geq 0 \) and \( \hat{\chi}^{-1} \geq 1 \) (i.e. periodic dynamics from a continuous dynamical system) are also aligned between the \( y \)-axes. The \( \hat{\chi}^{-1} \) curves in Figure 4.2c are distributed between the left and right \( y \)-axis and scaled for readability.

Observe in Figure 4.2b that \( \hat{\chi}^{-1} \) tracks the relative change in \( \lambda_1 \) and detects periodic windows far more effectively than \( h_{GNE} \) or \( h_{PE} \) and for smaller \( m \) (see Figure 3.2; entropy measures were computed for \( m = 14 \) in Chapter 3). Furthermore, \( \hat{\chi}^{-1} \) appears far less sensitive to period doubling bifurcations than all other measures considered thus far in this thesis. The most notable discrepancy between the estimates of \( \lambda_1 \) and \( \hat{\chi}^{-1} \) is that our new measure appears less effective at detecting some regimes
characterised by narrow band chaos. This can be observed as the system bifurcates out of the period-3 window at $\alpha \approx 0.41$. Figure 4.2c shows that $\hat{\chi}^{-1}$ is still very effective when the data is down-sampled, although the value of the measure is higher across the full domain of $\alpha$ than for the finely sampled time series which is very likely due to node aliasing. For example, $\hat{\chi}^{-1} \approx 2$ in the period-1 regime and $\hat{\chi}^{-1} \approx 3$ for period-4 time series. Node aliasing effects the network diameter because the alias edges effectively create shortcuts in the Markov chain. The addition of noise also appears to increase the baseline value and range of $\hat{\chi}^{-1}$ significantly. Despite these issues, $\hat{\chi}^{-1}$ still tracks the relative change in $\lambda_1$. Even with AWGN at SNR$_{dB} = 13.01$ dB some of the periodic windows are still detectable if one considers the relative values computed from the chaotic dynamics in the domain immediately surrounding the periodic window (i.e. see $\alpha \approx 0.4$).

4.3.2 The Muthuswamy-Chua system with quadratic nonlinearity (MC2)

The Muthuswamy-Chua system with quadratic nonlinearity is a set of equations that models the behaviour of a theoretical three element electronic circuit that was first proposed by Muthuswamy and Chua [2010]. It comprises a linear inductor, linear capacitor and an active nonlinear memristor (Chua [1971]; Chua and Kang [1976]) combined in series. The active nonlinear memristor is a theoretical circuit component that currently can only be physically implemented by multi-component circuits which emulate the differential equations that describe the memristor’s behaviour. The MC2 system is defined by the equations:

\[
\begin{align*}
\frac{dx}{dt} &= \frac{y}{C}, \\
\frac{dy}{dt} &= -\frac{1}{L}(x+yR(z)), \\
\frac{dz}{dt} &= -y + z(y - \rho),
\end{align*}
\]

where:

\[
R(z) = \zeta(z^2 - 1),
\]

and has parameters $\rho = 0.9$, $C = 1$, $L = 3$, and bifurcation parameter $\zeta \in [0.6, 1.6]$. We numerically integrate this system using a second-third order Runge-Kutta method to generate finely sampled
4.3. NUMERICAL INVESTIGATION

**Figure 4.2:** (a) The Rössler system bifurcation diagram, (b) the largest Lyapunov exponent $\lambda_1$ and inverse normalised ordinal network diameter $\hat{\chi}^{-1}$ for finely sampled data, and (c) $\hat{\chi}^{-1}$ for coarsely sampled data with various levels of AWGN.

$x$-component time series with length $N = 10^6$ (after transients have been removed) at a sampling period of $\Delta t = 0.01$. This corresponds to approximately 1020 samples for each period-1 cycle when $\zeta = 0.6$. To obtain coarsely sampled data, the time series are down-sampled by a factor of 10.

The ordinal networks are constructed with embedding lag $\tau = 320$ for finely sampled time series and equivalently $\tau = 32$ for coarsely sample time series. This lag corresponds approximately to the mean value of the first zero of the time series autocorrelation function over the ensemble of time series, which is about 317 samples. The embedding dimension is set at $m = 12$ for noiseless data and $m = 10$ for data with AWGN.
CHAPTER 4. CHARACTERISING TIME SERIES COMPLEXITY VIA INVERSE NORMALISED ORDINAL NETWORK DIAMETER

Figure 4.3: (a) A chaotic attractor generated by the MC2 system and (b) an ordinal network constructed from the x-component time series with embedding dimension $m = 12$.

An example of a chaotic MC2 attractor and its corresponding ordinal network are shown in Figure 4.3. Results for the computation of $\lambda_1$ and $\hat{\chi}^{-1}$ are shown in Figure 4.4. The new measure $\hat{\chi}^{-1}$ appears to be highly correlated with $\lambda_1$ for the MC2 toy data, and arguably outperforms the traditional invariant for periodic regimes. For example, the value of $\lambda_1$ for $\zeta \approx 1.08$ (immediately after the period-2 to period-4 bifurcation) is clearly erroneous, whereas $\hat{\chi}^{-1} \approx 1$ as expected for periodic dynamics. The method used to estimate the exponent (Sano and Sawada [1985]) hinges on the construction of a local linear flow map of the phase space dynamics by taking an $\epsilon$-neighbourhood of points around each state vector and observing the evolution of these points to approximate the tangent space. The error in $\lambda_1$ for $\zeta \approx 1.08$ has likely occurred because $\epsilon$ is too large, given that the period-4 trajectories will be very close following the period-doubling bifurcation. Our method does not require the selection of an $\epsilon$ or similar parameter, but rather depends on the topology of the ordinal partition in $m$-dimensional space to delineate the discrete states of the Markov chain. In this particular instance, $\hat{\chi}^{-1}$ is more effective than $\lambda_1$ in characterising the dynamics, and is further advantageous because its computation requires the selection of fewer method parameters.

The general trend that can be observed as noise is added to the down-sampled time series in Figure 4.4c is an increase in the minimum value and range of $\hat{\chi}^{-1}$, as per the results for the Rössler system. However, the measure still appears to be very effectively at detecting periodic windows despite AWGN with signal to noise ratios as low as $\text{SNR}_{\text{dB}} = 13.01 \, \text{dB}$ for the MC2 time series.

4.3.3 The Muthuswamy-Chua system with 4th-order polynomial nonlinearity (MC4)

Muthuswamy-Chua system with 4th-order polynomial non-linearity is a modification of the MC2 system that was first proposed by McCullough et al. [2013]. The equations which define the MC4 system are:

\[
\begin{align*}
\frac{dx}{dt} &= \frac{y}{C}, \\
\frac{dy}{dt} &= -\frac{1}{L}(x + yR(z)), \\
\frac{dz}{dt} &= -y + z(y^2 - \rho),
\end{align*}
\]  

(4.13)
4.3. NUMERICAL INVESTIGATION

Figure 4.4: (a) The MC2 system bifurcation diagram, (b) the largest Lyapunov exponent $\lambda_1$ and inverse normalised ordinal network diameter $\hat{\chi}^{-1}$ for finely sampled data, and (c) $\hat{\chi}^{-1}$ for coarsely sampled data with various levels of AWGN.

where:

$$R(z) = \left(\frac{1}{2}z^4 - \frac{3}{2}z^2 - \eta\right), \quad (4.14)$$

and has parameters $\rho = 0.9$, $C = 1$, $L = 3$, and bifurcation parameter $\eta \in [1.5, 4]$. The primary differences between MC2 and MC4 are that the latter uses a 4th-order double well polynomial memristance function and quadratic non-linearity in the second term of the $z$-component equation.

We numerically integrate this system using a second-third order Runga-Kutta method to generate finely sampled $x$-component time series with length $N = 10^6$ (after transients have been removed) at a sampling period of $\Delta t = 0.005$. This corresponds to approximately 1187 samples for each periodic cycle when $\eta = 2.0675$. To obtain coarsely sampled data, the time series are down-sampled by a factor of 10.
CHAPTER 4. CHARACTERISING TIME SERIES COMPLEXITY VIA INVERSE NORMALISED ORDINAL NETWORK DIAMETER

Figure 4.5: (a) A chaotic attractor generated by the MC4 system and (b) an ordinal network constructed from the $x$-component time series with embedding dimension $m = 8$.

The ordinal networks are constructed with embedding lag $\tau = 450$ for finely sampled time series and equivalently $\tau = 45$ for coarsely sampled time series. This lag corresponds approximately to the mean value of the first zero of the time series autocorrelation function over the ensemble of time series, which is about 475 samples. The embedding dimension is set at $m = 7$.

An example of a chaotic four-scroll attractor generated by the MC4 equations and its corresponding ordinal network are shown in Figure 4.5. Results for the computation of $\lambda_1$ and $\hat{\chi}^{-1}$ are shown in Figure 4.6. Again, our new measure tracks the relative change in $\lambda_1$ over the domain of $\eta$ even when data is down-sampled and corrupted with noise. Notable in this example is that the tracking is effective for $m = 7$ which is a relatively low embedding dimension.

4.3.4 The Chua circuit equations

The final system that we consider in this numerical investigation is the system of equations that models the Chua circuit (see Chua et al. [1992]):

\[
\begin{align*}
\frac{dx}{dt} &= \alpha(y - h(x)), \\
\frac{dy}{dt} &= x - y - z, \\
\frac{dz}{dt} &= -\beta y,
\end{align*}
\]

where:

\[ h(x) = m_1 x + \frac{m_0 - m_1}{2}(|x + 1| - |x - 1|), \]

and has parameters $\alpha = 15.6$, $\beta = 28 = 1$, $m_0 = -8/7$, $m_1 = -5/7$ and bifurcation parameter $\beta \in [25, 40]$. We numerically integrate this system using a second-third order Runge-Kutta method to generate finely sampled $x$-component time series with length $N = 10^6$ (after transients have been removed) at a sampling period of $\Delta t = 0.001$. This corresponds to approximately 1200 samples for
4.3. NUMERICAL INVESTIGATION

(b) The largest Lyapunov exponent $\lambda_1$ and inverse normalised ordinal network diameter $\hat{\chi}^{-1}$ for finely sampled data, and (c) $\hat{\chi}^{-1}$ for coarsely sampled data with various levels of AWGN.

Figure 4.6: (a) The MC4 system bifurcation diagram, (b) the largest Lyapunov exponent $\lambda_1$ and inverse normalised ordinal network diameter $\hat{\chi}^{-1}$ for finely sampled data, and (c) $\hat{\chi}^{-1}$ for coarsely sampled data with various levels of AWGN.

Each period-1 cycle when $\beta = 40$. To obtain coarsely sampled data, the time series are down-sampled by a factor of 10.

The ordinal networks are constructed with embedding lag $\tau = 440$ for finely sampled time series and equivalently $\tau = 44$ for coarsely sample time series. The embedding dimension is set at $m = 7$.

An example of a chaotic attractor generated by the Chua equations and its corresponding ordinal network are shown in Figure 4.7. Results for the computation of $\lambda_1$ and $\hat{\chi}^{-1}$ are shown in Figure 4.8. Our new measure is effective at tracking changes in dynamics for the Chua system time series but is evidently less stable than $\lambda_1$ for chaotic dynamics where $\beta \lessapprox 32$. The effect of AWGN is far more pronounced and irregular for this system, as can be observed in Figure 4.8c. The minimum value
and range of $\hat{\chi}^{-1}$ both increase as they did for the other systems, however, for $30 < \beta < 32$ and near the period doubling bifurcation at $\beta \approx 35.5$, even low levels of AWGN cause a sharp increase in the measure. It is possible that for these values of $\beta$, the topology of the attractor makes $\hat{\chi}^{-1}$ less robust to noise. For example, the method would be very sensitive to noise if there exist embedded states vectors from a noiseless time series which are close in terms of Euclidean distance, but opposite phase in terms of the period of the dynamics. In such an attractor, even a small amount of noise may translate to an edge in the ordinal network that connects two nodes which would otherwise be separated by a much longer geodesic path, and hence may greatly increase the rate of Markov mixing in the network model.

![Figure 4.7:](a) A chaotic attractor generated by the Chua system and (b) an ordinal network constructed from the $x$-component time series with embedding dimension $m = 10$.]

### 4.4 Application: Epileptic seizure onset detection and pre-detection from electroencephalograms

In this section we present results of an investigation of electroencephalogram (EEG) time series using ordinal networks and our novel measure $\hat{\chi}^{-1}$ to test the applicability of our methods for the analysis of complex, noisy, multivariate, spatio-temporal data that has been recorded in a clinical setting. Our primary aim is to examine the extent to which $\hat{\chi}^{-1}$ can characterise (and hence discriminate) the different dynamical states associated with normal brain function and an episode of an epileptic seizure. The secondary aim of this section is to attempt to identify a change in the brain dynamics prior to the annotated seizure onset time, thereby evidencing the possibility of short time pre-detection of epileptic seizures.

#### 4.4.1 Data and background

We are using data from the CHB-MIT Scalp EEG database on PhysioNet (Goldberger et al. [2000]) which was first reported and investigated by Shoeb [2009] where machine learning techniques were used to implement a patient specific seizure onset detection scheme. The data comprises multi-channel EEG recordings from 23 subjects with epilepsy that are grouped into 24 cases. For the purpose of
4.4. APPLICATION: EPILEPTIC SEIZURE ONSET DETECTION AND PRE-DETECTION FROM ELECTROENCEPHALOGRAMS

Figure 4.8: (a) The Chua system bifurcation diagram, (b) the largest Lyapunov exponent $\lambda_1$ and inverse normalised ordinal network diameter $\hat{\chi}^{-1}$ for finely sampled data, and (c) $\hat{\chi}^{-1}$ for coarsely sampled data with various levels of AWGN.

In this study we treat each of the 24 cases as a unique subject. The data was sampled at 256Hz with 16 bit resolution and the majority of records have 23 channels of data. Note that we incorporate any additional data in our analysis where records have more channels available. Each channel corresponds to a voltage measurement between spatially distributed electrodes placed on the subject’s scalp as per the standard International 10-20 system (Klem et al. [1999]). Records for each patient span up to several days and contain multiple seizure events. The database includes annotations for the start and end times of each seizure event as determined by an expert.

In the original study, Shoeb [2009] trained a binary nonlinear support vector machine (SVM) for each subject using two or more seizure events and then applied the SVM to the remainder of a subject’s
records to detect the onset of any further seizures. The nonlinear SVM used radial basis functions to classify the feature vectors. The feature vectors were constructed as follows. For a given channel the authors compute a histogram of the power spectral density, a linear property of the time series, with $M$ bins for a short time window ending at time index $t$. This $M$-long vector is computed for all $K$ channels and concatenated to form an $M \times K$-long vector. This vector is then concatenated along the time dimension with all those in a second short time window length $T$ preceding time index $t$ to form the feature vector which has $M \times K \times T$ elements. These feature vectors can therefore be understood as encoding the spatial distribution over different channels of the EEG of the short time evolution of the power spectral density.

### 4.4.2 Seizure onset detection

**Objectives and methodology**

As previously stated, our primary aim is to test the effectiveness of $\hat{\chi}^{-1}$ as a measure which discriminates between EEG time series for normal brain activity and for epileptic seizures, which might then enable computerised seizure onset detection. We limit the scope of our investigation to an in-sample analysis. That is, we do not attempt to construct a detector and assess its performance in detecting out-of-sample events as per Shoeb [2009]. Instead, we train a simple binary classifier for each annotated seizure event and the data that precedes it, and then assess the performance of the classifier against the annotations. If the binary classifier flags seizure onset within a short time of the annotated onset time, and has a low rate of false positives, then it can be inferred that $\hat{\chi}^{-1}$ is an effective measure for discriminating between the dynamics corresponding to a normal brain state and the seizure state.

The methodology for our analysis is detailed as follows. For a given seizure event from a given subject we extract from the database all channels of data between the annotated start and end times of the seizure, henceforth called the ictal phase, and 10 minutes of data leading up to the onset of the ictal phase, henceforth called the pre-ictal phase. Sometimes it is not possible to extract a full 10 minutes of pre-ictal phase data because two seizures occur in close succession, or because there is not enough data at the beginning of a record before the onset of the ictal phase. If this is the case then we take pre-ictal phase data starting from the end of the previous seizure or the beginning of the record. No pre-processing is applied to the data.

We perform a sliding window analysis for each channel (CH1, CH2, ..., CH23) separately as illustrated in Figure 4.9. The sliding window has length of $L = 5$ seconds or 1280 samples and steps along the time series in increments of $T_{\text{step}} = 0.25$ seconds or 64 samples. Note that the window is indexed at its leading edge (i.e. the window comprises data for time indices $\{t - L, t - L + T_{\text{step}}, t - L + 2T_{\text{step}}, \ldots, t\}$) so that $\hat{\chi}^{-1}$ is only computed with respect to past data. For each window we construct a weighted ordinal network given by the adjacency matrix $A$ with embedding parameters $m = 3$ and $\tau = 2$. A small embedding dimension is selected for three reasons: to enable rapid computation of $\hat{\chi}^{-1}$ over the large dataset by limiting the network size; to ensure good sampling of the distribution of ordinal patterns and transitions given that each window only contains 1280 data points; and to define a coarse partition so that the network is more robust against noise. We did not thoroughly investigate the parameter space for $\tau$ but report that the selected parameter value $\tau = 2$ performed notably better than $\tau = 1$ and $\tau = 5$. We then compute the modified ordinal network $A^*$ based on Equation 4.8 and...
Figure 4.9: A flowchart depicting the application of ordinal network analysis to a single channel of EEG data: The data is shown in the top panel. The onset and duration of the ictal phase are marked respectively by the vertical dotted orange line and highlighted range that follows. These are known a priori from expert annotations in the data. The green bar denotes a 5 second sliding window that is used to extract a short segment of data which is mapped to an ordinal network. The inverse normalised network diameter \( \hat{\chi}^{-1} \) is then computed to estimate the mixing rate of that particular window. This process is repeated for all windows along the data to produce a new time series \( \hat{\chi}^{-1}_{\text{CH1}}(t) \) which enables us to observe changes in complexity along the time series.
then the inverse normalised network diameter with respect to the time index of the window $t$, given by $\hat{\chi}^{-1}_{\text{channel}}(t)$. Observe in the example shown in Figure 4.9 that the measure remains relatively constant until the onset of the ictal phase, at which time it increases rapidly. However, given the complex nature of neural dynamics we do not expect (and nor do we observe) a uniform spatial response in the measure at the onset of the ictal phase, and the response that we do observe is certainly not consistent between patients, hence a simple threshold is not sufficient to classify the onset.

Therefore, we define the feature vector over all channels for the window at time index $t$ as:

$$X(t) = \{ \hat{\chi}^{-1}_{\text{CH1}}(t), \hat{\chi}^{-1}_{\text{CH2}}(t), \hat{\chi}^{-1}_{\text{CH3}}(t), \ldots, \hat{\chi}^{-1}_{\text{CH23}}(t) \},$$

(4.17)

where each element can be understood as a scalar that quantifies a nonlinear property of the dynamics from single channel of the EEG over a short time window. By way of comparison, the feature vectors defined by Shoeb [2009] were larger by two multiplicative factors $K$ and $T$ and were based on a linear measure of the dynamics. We then construct a binary classification model using a linear SVM as implemented in the MATLAB function `fitclinear`. This function takes a set of feature vectors and a corresponding set of class labels as its input. Optimisation by stochastic gradient descent is used to find a linear projection of the feature vectors onto a 1-dimensional line such that the two classes are separated by the widest possible margin. While it may be possible to construct a better classifier using nonlinear support vector machines or other machine learning models, we specifically choose this simple linear classifier to test the extent to which the ordinal network method and our new measure $\hat{\chi}^{-1}$ can process the EEG data such that the feature vectors are linearly separable. In other words, we want the ordinal network to perform the majority of the work. Each feature vector is labelled as belonging to either the pre-ictal phase (label 0) or the ictal phase (label 1). The SVM is trained using the set of feature vectors and corresponding labels. The learning algorithm is weighted such that the cost of a false positive (i.e. misclassifying a feature vector from the pre-ictal phase as belonging to the ictal phase) is 75 times greater than for a false negative. The trained model is then used to reclassify the same set of feature vectors as either pre-seizure (class 0) or ictal phase (class 1). We consider onset detection successful if at least one feature vector from the ictal phase is classified as belonging to the ictal phase class within 30 seconds of seizure onset.

Figure 4.10 shows a visualisation of the feature vectors and the resulting SVM classification scores and classifications for a single seizure event. Observe that the plot of the feature vectors shows a distinct change in the spatial and temporal nature of the dynamics during the ictal phase. In this example the SVM successfully detects the seizure onset with a small detection delay time $t_{\text{delay}}$.

Results

We perform the analysis for all seizure 198 seizure events in the database. The onset detection success rate and count of false positives per subject are shown in Figure 4.11. The mean success rate is 87% with a total of 10 false positives which is equivalent to less than 8 false positives per 24 hours. Note that the low success rate for subject 12 may be attributable to especially severe epilepsy. Subject 12 suffered 40 seizures during the period of observation which was twice as many as the next highest count and 3 times higher than the mean seizure count per subject of 13.2. If we consider this case an outlier then the mean success rate is 96% with a total of 9 false positives. Figure 4.12 shows the onset
4.4. APPLICATION: EPILEPTIC SEIZURE ONSET DETECTION AND PRE-DETECTION FROM ELECTROENCEPHALOGRAMS

![Figure 4.10: Seizure onset detection using a binary support vector machine (SVM) and ordinal network analysis for multi-channel EEG data: The top panel depicts all 23 channels of \( \hat{\chi}^{-1}(t) \) time series that have been computed using a sliding window analysis as described in Figure 4.9 for a single seizure event from a single subject. The lower panels show the in-sample SVM scores and subsequent classification of the feature vectors \( X(t) = \{ \hat{\chi}_{CH1}^{-1}(t), \hat{\chi}_{CH2}^{-1}(t), \hat{\chi}_{CH3}^{-1}(t), \ldots, \hat{\chi}_{CH23}^{-1}(t) \} \). The annotated onset and duration of the ictal phase are marked in the lower panels respectively by the vertical dotted orange line and highlighted range that follows.]

Detection delay \( t_{\text{delay}} \) for all 198 seizure events with a mean of 3.77 seconds and median of 2.75 seconds. For reference, Shoeb [2009] reported an out-of-sample success rate of 96%, a rate of 2 false positives per 24 hour period and a median onset detection delay of 3 seconds for 163 test seizures using their linear SVM detector.

Our results show that \( \hat{\chi}^{-1} \) effectively compresses the multivariate time series into a feature vector that can be linearly separated to classify the pre-ictal and ictal dynamics with a high success rate. The primary purpose of this investigation has been a proof of concept that our proposed method is applicable to real data. When applying our method in an experimental or clinical study it would be necessary to divide the data into a training and test data set, or alternatively implement an unsupervised learning algorithm or a scheme for change point detection. It also must be restated that we deliberately chose to use a linear SVM and did not perform any parameter optimisation (beyond testing several values for the misclassification cost) to test the ordinal network method with minimal reliance on the SVM.
Figure 4.11: (a) The success rate of onset classification, and (b) the number of false positive classifications per subject.

Figure 4.12: Histogram of the onset classification delay $t_{\text{delay}}$ for all seizure events in the data set where onset classification was successful.
4.4. APPLICATION: EPILEPTIC SEIZURE ONSET DETECTION AND PRE-DETECTION FROM ELECTROENCEPHALOGRAMS

4.4.3 Identification of pre-seizure state

Objectives and methodology

In this section we turn our attention to the challenge of pre-detecting epileptic seizures. Our specific objective is to find evidence of a state change in the spatio-temporal EEG time series dynamics prior to the seizure. We proceed with a simple test as follows. For a given seizure event we extract the pre-ictal time series from the database and compute the set of feature vectors as per Section 4.4.2. Note that we do not consider the time series for the ictal phase. We then define a pre-detection window immediately prior to the ictal phase. The length of the pre-detection window is arbitrarily chosen to be $T_{\text{pre-window}} = 20$ seconds. Each feature vector is labelled as belonging to either the pre-detection window (label 1) or the remainder of the pre-ictal phase (label 0). The SVM is trained using the set of feature vectors and corresponding labels. The learning algorithm is weighted such that the cost of a false positive (i.e. misclassifying a feature vector from the pre-ictal phase as belonging to the pre-detecting window) is 65 times greater than for a false negative. The trained model is then used to reclassify the same set of feature vectors as either pre-seizure (class 0) or within the pre-detecting window (class 1). We consider the identification of a state change in the pre-detection window successful if at least one feature vector from the pre-detection window is classified as belonging to the pre-detection window class.

Figure 4.13 shows a visualisation of the feature vectors and the resulting SVM classification scores and classifications for a single seizure event. In this example the SVM successfully detects a change in the system state within the pre-detection window that has not been observed at any earlier point in the time series. It is critical to understand that successful detection should not be directly interpreted as a pre-detection of seizure onset. This is due to the fact that the test we describe above begins with the assumption that a state change occurs exactly 20 seconds before the onset of the ictal phase. The model is then trained based on this assumption and then tested in-sample to determine whether or not it was possible to linearly separate the feature vectors. However, if the in-sample classification implies a state change within the pre-detection window, this is evidence that short time pre-detection of epileptic seizures may be possible using $\chi^{-1}$.

Results

Success rates and false positive counts for the pre-detection test as applied to all 198 records in the database are shown in Figure 4.14. The mean success rate is 45.95% with a total of 11 false positives which is equivalent to less than 10 false positives per 24 hours. Figure 4.15 shows the pre-detection time $t_{\text{pre-detect}}$ for all 198 seizure events with a mean of 7.18 seconds and median of 17.25 seconds before seizure onset.

These results show that our method detects a state change during the pre-detection window for almost half of the seizure events. Moreover, the method is far more likely to detect this state change than to produce a false positive. This is a promising indication that short-time pre-detection of seizure onset may be possible using $\chi^{-1}$. The primary limiting factor in this investigation is the assumption underpinning our hypothesis that a state change occurs at some fixed time $T_{\text{pre-window}}$ before onset for all seizures and subjects. If a pre-seizure state does manifest in a given record but the change to this
Figure 4.13: Seizure onset pre-detection using a binary support vector machine (SVM) and ordinal network analysis for multi-channel EEG data: The top panel depicts all 23 channels of $\hat{\chi}^{-1}(t)$ time series that have been computed using a sliding window analysis as described in Figure 4.9 for a single seizure event from a single subject. The annotated onset and duration of the ictal phase are marked in the lower panels respectively by the vertical dotted orange line and highlighted range that follows. We assume the existence of a state change in the 20 seconds immediately prior to seizure onset (highlighted blue) and attempt to train a SVM to discriminate between this window and the remainder of the pre-seizure data (discarding data from the ictal phase). The lower panels shows the in-sample SVM scores and subsequent classification of the feature vectors $X(t)$.

state occurs at some time before or after $T_{\text{pre-window}}$, which it almost certainly will in the absence of prior knowledge of the true state change time (see the apparent change in dynamics as characterised by the feature vectors at $t \approx 1800$ in Figure 4.13), then the class labels which we have assumed for some of the feature vectors will not correspond to the true system state. This will hamper the SVM learner in finding a hyperplane to separate the feature vectors and detect the true state change, and likely result in either a detection delay or false positives, or in the worst case, a complete failure to detect the existing pre-seizure state. Hence, we believe that a far higher success rate could be achieved if our method was implemented in conjunction with a scheme for change point detection or unsupervised learning.
4.4. APPLICATION: EPILEPTIC SEIZURE ONSET DETECTION AND PRE-DETECTION FROM ELECTROENCEPHALOGRAMS

Figure 4.14: (a) The success rate of onset pre-detection, and (b) the number of false positive classifications per subject.

Figure 4.15: Histogram of the pre-detection time $t_{\text{pre-detect}}$ for all seizure events in the data set where pre-detection was successful. Time $t_{\text{pre-detect}} = 0$ corresponds to the annotated seizure onset time and lower values indicate earlier pre-detection. For example, $t_{\text{pre-detect}} = -18$ means that the a distinct pre-seizure feature vector was detected 18 seconds before the annotated onset time.
4.5 Chapter summary

In this chapter we defined a new measure called the inverse normalised network diameter which is a global measure of ordinal network topology. We have shown that this measure effectively characterises the complexity of the dynamics in noisy chaotic time series and can be used to discriminate between different physiological states from multivariate biomedical signal data recorded in a clinical setting when used in conjunction with simple machine learning techniques.

The theoretical basis for our measure is that the network diameter is a lower bound to the Markov chain mixing time. In addition to existing literature pertaining to this relationship, we demonstrated that the time taken for an irreducible and aperiodic Markov chain to converge within a total variation distance equal to the smallest probability in the stationary distribution is bounded from below by the network diameter. Furthermore, we have shown that the only condition necessary to guarantee that an ordinal network is irreducible is that there must exist a finite length directed path from the node corresponding to the final symbol in the symbolic dynamics, to the node corresponding to the first symbol. This path will exist for ordinal networks constructed from ergodic time series if the data is sufficiently long that a recurrence of the initial state is observed. In practice, where the length of the data is fixed and insufficiently sampled or corrupted by noise, it is usually possible to truncate the start or end of the data to remove transient states. Alternatively, a smaller embedding dimension can be used to make the ordinal partition coarser and increase the likelihood of a recurrence of the initial state.

The newly proposed measure, the inverse normalised network diameter, is computed on a modified adjacency matrix where edges are equal to the average time spent in the previous state and a single time step to account for the state transition. The average time spent in a given node or state is encoded in the self edge weights. This information is lost if diameter is computed on the original ordinal network because a geodesic can never contain a self edge. Therefore, computing weighted diameter on the modified network results in a measure which has units of time and incorporates in the information contained in the edge weights. This also provides less conservative estimates of the Markov mixing time under the assumption of a strong deterministic component in the dynamics and slow evolution of states relative to the sampling period of the data. We completed a comparative numerical investigation of the inverse normalised network diameter and estimates of the largest Lyapunov exponent using discretely sampled data from four different continuous chaotic systems. Our results show that the inverse normalised network diameter could reliably track the relative change in the largest Lyapunov exponent through various bifurcations and in the presence of additive white Gaussian noise with signal to noise ratios as low as $\text{SNR}_{\text{dB}} = 13.01 \text{dB}$. The inverse normalised network diameter appears to perform far better in this task than the entropy based measures and simple network statistics which have been examined in the previous chapters. This result supports our postulate that our new measure is related to the largest Lyapunov exponent for deterministic chaotic data, and indicates that the ordinal network encodes a useful stochastic approximation of the chaotic mixing process.

We applied the inverse normalised network diameter to analyse a large database of multivariate EEG time series that were recorded in a clinical setting from patients with epilepsy. Specifically, we constructed feature vectors from the new measure as computed in a sliding window analysis. The resulting feature vectors therefore characterised spatial, temporal and non-linear information and have
length equal to the number of channels in the data. Using a binary linear SVM we demonstrated that
the features vectors corresponding to the ictal and pre-ictal states, respectively, are linearly separable
with a high success rate, low detection delay and relatively few false positives. In almost half of the
cases we also identify a pre-seizure state change with low rate of false positives. This result evidences
the possibility of devising a scheme for pre-detection using the inverse normalised network diameter.
Our investigations were strictly limited to in sample classification. That is, we used the SVM with
minimal parameter optimisation to test the extent to which our measure could discriminate between
different states in the spatio-temporal dynamics, and in this capacity the measure was very effective.
Chapter 5

Generating surrogate time series from ordinal network

ABSTRACT

In Chapters 2 though 4 we have shown that the structural properties of an ordinal network can be used to quantify the complexity of the time series from which they have been constructed. In this chapter we investigate the extent to which ordinal networks encode the dynamics of nonlinear time series in the structure of the networks from a different perspective. That is, we study the network model’s capacity to generate new time series with similar dynamical properties to the original data. To do this we take constrained random walks on the network to regenerate new symbolic dynamics and use a data reassignment procedure to produce surrogate time series. We then compute invariant measures and recurrence properties to compare the original time series with the regenerated surrogates. The remainder of the chapter is structured as follows. In Section 5.2 we define the constrained random walk algorithm and a data reassignment procedure for regenerating surrogates. In Section 5.3 we present a qualitative analysis of the random walk surrogate time series and then use order recurrence plots, recurrence quantification analysis, and estimates of the largest Lyapunov exponent in a comparative study of the properties of the surrogates. Finally, in Section 5.4, we investigate the out-of-sample predictive properties of ordinal networks for low dimensional chaotic time series.

5.1 Generating time series from network models

To recapitulate broadly the primary concept underpinning this thesis: network based methods of nonlinear time series analysis amount to constructing models of dynamical systems from univariate or, in some cases, multivariate time series, and then computing various statistics on the model to quantify the properties of the model. For the purpose of demonstrating the merit of this idea and developing it into a framework for practical application, we seek to test the hypothesis that the model properties are dependant on the dynamical properties of interest in the system from which the time series has been observed. A key characteristic of these models is that they require relatively few user selected parameters because the overwhelming majority of the information in the model is a transformation of
CHAPTER 5. GENERATING SURROGATE TIME SERIES FROM ORDINAL NETWORK

the time series. Given that our aim is to establish the connection between the time series dynamics and the model, this raises two related but distinct questions. Firstly, is it possible to perform an inverse transformation and reconstruct the original data from the network? Secondly, if it is not possible to perform an inverse transformation, is the network model an analogue of the original system?

The possibility of performing an inverse transform hinges on how much information from the time series has been encoded in the network. Strictly speaking, an inverse transform is only possible if the time series and the network are topologically conjugate. Thiel et al. [2004] devised an algorithm to reconstruct time series from binary recurrence plots\(^1\) provided that the connectivity threshold parameter (the Euclidean distance between embedding vectors in reconstructed state space) is set such that the resulting recurrence plot is a single fully connected component. Their numerical investigation showed that the inverse transformation produced a time series which was the same length as the original data and appeared to be topologically equivalent. Hirata et al. [2008] proposed an alternative algorithm whereby the binary recurrence matrix is mapped to a weighted recurrence network. From this network they computed the inter-node distances and used multidimensional scaling to map these back to a univariate time series. Estimates of the largest Lyapunov exponent were almost identical for the original and reconstructed time series, indicating the preservation of invariant topological properties in the examples which they presented. This method was later applied by Hirata et al. [2016] to reconstruct 3 dimensional images of chromosome structures.

However, both of the aforementioned inverse transformations operate on proximity network models of time series, whereas ordinal networks are Markov model approximations. If the time series is deterministic, the time evolution of an ordinal network model will be governed by a fundamentally different mechanism than that of the original system and, hence, the model is not topologically conjugate to the time series. Even for systems where it has been shown that the ordinal partition is generating (Amigó and Keller [2013]; Keller and Sinn [2010]), which implies that the symbolic dynamics \(s\) are topologically conjugate, the map from \(s\) to the network only explicitly encodes the distribution of symbols and 2-letter words from \(s\). Therefore, while it is possible for the Markov process to realise a true symbolic trajectory from the original system, information about deterministic correlations for trajectories with length greater than 2 cannot be extracted from an ordinal network in the general case.

This leads us back to the question of whether or not an ordinal network is an analogue of the original system. We use the term analogue to mean a system which generates new time series with properties similar to the original data used to construct the model, where similarity is assessed by computation of appropriate measures between the original data and the new time series. For example, if the original data was observed from a low dimensional chaotic system, one can compare features of recurrence plots or estimate invariant measures. This can be understood as a form of in-sample assessment of the model but can easily be extended for out-of-sample comparison by making use of separate training and test data (as we do in Section 5.4). In the remainder of this chapter, we shall proceed with this general methodology to investigate the extent to which ordinal networks encode the dynamics of discrete sampled time series from continuous chaotic systems in the structure of the networks.

\(^1\)Recall from Chapter 1 that a recurrence plot can be interpreted as the adjacency matrix of a network.
5.2 Methodology

5.2.1 Preliminaries

Consider a time series \( x = \{x_n\}_{n=1}^N \) embedded in \( m \)-dimensional space with lag \( \tau \) to form the set of embedded state vectors \( \{z_n\}_{n=1}^{N-(m-1)\tau} \) where \( z_n = (x_n, x_{n+\tau}, x_{n+2\tau}, \ldots, x_{n+(m-1)\tau}) \). An ordinal partition is then applied to the time series to construct the symbolic dynamics \( s = \{s_n\}_{n=1}^{N-(m-1)\tau} \) for \( s_n \in S \). This is subsequently mapped to the network given by the adjacency matrix \( A \) with directed edges \( a_{i,j} \). Edges have weight equal to the frequency of forward time transitions for any given pair of symbolic states in \( s \) as represented by nodes \( i \) and \( j \). Refer to Chapter 3 Section 3.1 for a complete description of the method for constructing a weighted ordinal network.

5.2.2 Random walks on ordinal networks

As discussed in Chapter 3, the information encoded in an ordinal network can be used to build a first-order Markov chain of \( s \) where \( p_{i,j} \) is the conditional probability of a transition from node \( i \) to node \( j \) as defined in Equation 3.3. Using this model, one can select a random initial node on the network and perform a Markovian random walk to generate a new symbolic sequence \( \hat{s} \) of length \( L \). However, when \( \tau > 1 \) this simple model does not strictly adhere to the transitional properties of ordinal symbolic dynamics. To explain this we shall first consider the case for \( \tau = 1 \). The temporally adjacent ordinal symbols \( s_n \) and \( s_{n+1} \) correspond to the embedding vectors \( z_n = (x_n, x_{n+1}, \ldots, x_{n+(m-1)}) \) and \( z_{n+1} = (x_{n+1}, x_{n+2}, \ldots, x_{n+(m)}) \) respectively, which have \( (m-1) \) overlapping elements from \( x \). Therefore, it must hold that \( \Lambda(s_n(2,\ldots,m)) = \Lambda(s_{n+1}(1,\ldots,m-1)) \) where \( s_k(i,\ldots,j) = (\pi_1,\ldots,\pi_{m-1}) \) (e.g. the ordered set of elements from \( s_n \) for indices \( i,\ldots,j \)) and the operator \( \Lambda \) gives the amplitude rank of the elements in a set (or equivalently the rank ordering). For example, the symbol \( \{3,1,2\} \) can only be followed by \( \{1,2,3\}, \{1,3,2\} \) or \( \{2,3,1\} \). We call these the allowable ordinal transitions. The procedure by which edges are allocated in an ordinal network guarantees that a Markovian random walk can only ever produce allowable ordinal transitions when \( \tau = 1 \).

Now consider the case for \( \tau > 1 \). Temporally adjacent symbols \( s_n \) and \( s_{n+1} \) correspond to the embedding vectors \( z_n = (x_n, x_{n+\tau}, x_{n+2\tau}, \ldots, x_{n+(m-1)\tau}) \) and \( z_{n+1} = (x_{n+1}, x_{n+1+\tau}, x_{n+1+2\tau}, \ldots, x_{n+1+(m-1)\tau}) \) which contain no overlapping elements from \( x \). In fact, the ordinal symbol \( s_n \) places constraint not on its temporal neighbour \( s_{n+1} \), but rather on the symbol \( s_{n+\tau} \). To ensure that this constraint holds for a random walk on the network, the walk algorithm must look back \( (\tau-1) \) steps from the current node to determine which of the connected nodes, if any, constitute an allowable ordinal transition. For simplicity, assume that the time index of a random walk \( \hat{s} \) at the current node \( i \) is \( (n-1) \). The transition from node \( i \) to any node \( j \) is an allowable ordinal transition with respect to an arbitrary ordinal symbolic dynamics with embedding dimension \( m \) if and only if \( \Lambda(\hat{s}_{n-\tau}(2,\ldots,m)) = \Lambda(\hat{s}_{n}(1,\ldots,m-1)) \).
CHAPTER 5. GENERATING SURROGATE TIME SERIES FROM ORDINAL NETWORK

It should be apparent that an initial trajectory \( \hat{s}_{initial} \in s \) of length \( L_{init} = \tau \) must be specified for a random walk which incorporates a memory of the past \( \tau \) states.

We now define our algorithm which generates a random walk \( \hat{s} \) of length \( L \) steps on the network \( A \) such that allowable ordinal transitions are enforced. For the remainder of the chapter we shall refer to this algorithm as the constrained random walk. The algorithm starts by selecting an initial trajectory \( \hat{s}(1, \ldots, \tau) \) randomly from \( s \). Note that here and in the pseudo-code to follow, parentheses immediately following an ordered set denote an index or indices to an element or an ordered set of elements in that set respectively and hence, for example, \( \hat{s}(b) = \hat{s}_b \) and \( s(b, \ldots, c) = \{s_b, \ldots, s_c\} \). The step index of the walk is stored in the pointer variable \( PTR \) which is initialised to \( \tau \). The algorithm then walks stepwise on the network from node \( i \) at \( PTR \) to node \( j \) at \( PTR + 1 \) based on the transitional probabilities as given by the edges \( a_{i,j} \) and conditional on the set of allowable transitions from \( \hat{s}(PTR - \tau + 1) \). If there are no allowable edges \( a_{i,j} \) then the algorithm steps backwards. A simple blocking procedure is implemented using the array variable \( BLK_n \) such that the walk will trace backwards from a proverbial dead-end state to a node from which an allowable path can be found, and then step forward again without the possibility of walking back into the same dead end until \( \hat{s} \) reaches \( L \) steps. If a path cannot be found then the algorithm will break and exit\(^2\). The pseudo-code for our algorithm is as follows.

\begin{verbatim}
1:  \( r \leftarrow \) random integer \( \in \{1, \ldots, |s| - \tau + 1\} \)
2:  \( \hat{s}(1, \ldots, \tau) \leftarrow s(r, \ldots, r + \tau - 1) \)
3:  \( PTR \leftarrow \tau \)
4:  \textbf{while} \( PTR < L \) \textbf{do}
5:  \( i \leftarrow \hat{s}(PTR) \)
6:  \( u \leftarrow \{\} \)
7:  \( BLK_n \leftarrow \{\}, \text{for all } n \in \{PTR + 1, \ldots, L\} \)
8:  \textbf{for} \( j : A_{i,j} > 0 \) \textbf{do}
9:     \textbf{if} \( j \) is an allowable ordinal transition from \( i \) given \( \hat{s}(PTR - \tau + 1) \) \textbf{then}
10:        \( u \leftarrow u \cup \{j\} \)
11:     \textbf{end if}
12: \textbf{end for}
13:  \( u \leftarrow u \cap BLK_{\text{PTR}} \)
14:  \textbf{if} \( u \neq \{\} \) \textbf{then}
15:     \( j \leftarrow \) random element \( j \in u \) given \( \Pr(j) = \frac{a_{j,i}}{\sum_{k \in u} a_{i,k}} \)
16:     \( PTR \leftarrow PTR + 1 \)
17:     \( \hat{s}(PTR) \leftarrow j \)
18:  \textbf{else}
19:     \( PTR \leftarrow PTR - 1 \)
20:     \( BLK_{\text{PTR}} \leftarrow BLK_{\text{PTR}} \cup \{i\} \)
21:  \textbf{end if}
22: \textbf{if} \( PTR < \tau \) \textbf{then}
\end{verbatim}

\(^2\)This can only occur when the final symbol or sequence of symbols in the symbolic dynamics has not occurred at any prior time in dynamics, which is very unlikely in practice as long as the size of the network is much greater than the length of the time series.
5.3. IN-SAMPLE ANALYSIS OF ORDINAL SURROGATE TIME SERIES

23: \textbf{break} and \textbf{exit} \{If \( PTR < \tau \) this implies that there is no random walk path of length \( N \) from the initial symbolic trajectory starting at \( r \). The user should reattempt with a different initial trajectory. See also Footnote 2 on page 80.\}

24: \textbf{end if}

25: \textbf{end while}

26: \textbf{output} ← \( \hat{s} \)

5.2.3 Regenerating time series from random walks

To regenerate a time series from the random walk symbolic dynamics \( \hat{s} \), we randomly reassign data from the original time series based on the ordinal symbolic map as follows. For all \( n \in 1, \ldots, L \) we find all embedding vectors \( z_{k} \) from the original time series \( x \) which map to symbol \( \hat{s}_{n} \). We then select one of these \( z_{k} \) at random and assign its first element to time index \( n \) such that \( \hat{x}_{n} = x_{k} \) where \( \hat{x} = \{\hat{x}_{n}\}_{n=1}^{L} \) is the regenerated time series. Assignment of \( x_{k} \) from \( z_{k} \) is performed with replacement, which is to say that \( x_{k} \) can be assigned multiple times in \( \hat{x} \) when it may only appear once in \( x \). The regenerated data \( \hat{x} \) can be considered as a type of surrogate time series (Hou et al. [2015]; Schreiber and Schmitz [2000]) for \( x \) based on the ordinal network model.

5.3 In-sample analysis of ordinal surrogate time series

In this section we investigate surrogate time series that are regenerated from ordinal networks with a time lag parameter \( \tau > 1 \) using the constrained random walk algorithm in comparison with the data used to construct the ordinal network model. This investigation is executed in three subsections, each with a distinct purpose. Firstly, in Section 5.3.1, we compare the qualitative appearance of time series and reconstructed attractors. Secondly, in Section 5.3.2 we compare the original ordinal symbolic dynamics \( s \) with random walk symbolic dynamics \( \hat{s} \) using order recurrence plots and order recurrence quantification analysis. We do this to determine how well ordinal networks and constrained random walks can preserve the temporal structure of the symbolic dynamics. Thirdly, in Section 5.3.2 we investigate how effectively invariant properties are preserved through the complete process of mapping a time series to an ordinal network then regenerating a surrogate. Specifically, we estimate the largest Lyapunov exponent for the original time series \( x \) and a set of surrogate time series \( \hat{x} \).

In each section we perform the same set of tests for two different time series from low dimensional chaotic systems. Both time series were generated by numerically solving their respective systems using a fourth-fifth order Runge-Kutta algorithm. Transients have been removed. The first is an \( x \)-component time series from the Lorenz system as governed by the equations:

\[
\begin{align*}
\frac{dx}{dt} &= \sigma(y - x), \\
\frac{dy}{dt} &= x(\rho - z) - y, \\
\frac{dz}{dt} &= xy - \beta z.
\end{align*}
\] (5.1)
This system is in the archetypal chaotic regime with parameters $\rho = 28$, $\sigma = 10$ and $\beta = 8/3$. The time series comprises 10000 uniformly sampled observations with sampling period $\Delta t = 0.01$ which results in approximately 100 points per mean cycle. The second time series is the $x$-component observed from the Rössler system as governed by Equation 2.4 with parameters $\alpha = 0.43$, $\beta = 2$ and $\gamma = 4$ such that the system is in fully developed chaos. This time series comprises 10000 uniformly sampled observations with sampling period $\Delta t = 0.1$ (approximately 60 points per mean cycle).

5.3.1 Qualitative comparison

Figures 5.1 and 5.2 show, for the Lorenz and Rössler time series respectively, the complete process of taking time series data, shown as time delay reconstruction of the attractor in this instance, then mapping that data to an ordinal network and, finally, regenerating a surrogate time series using a constrained random walk and data reassignment. Network visualisations were produced using a spring-electrical embedding algorithm in Wolfram Mathematica. It can be observed that the ordinal network encodes topological features of the attractors including the two wings and the separatrix of the Lorenz attractor, as shown in Figure 5.1b, and the manner in which the band of the Rössler attractor stretches and then folds back onto itself, as shown in Figure 5.2b. Visual inspection of the reconstructed attractors for the surrogate time series in Figures 5.1c and 5.2c suggests that ordinal network mapping and the random walk constitutes a good stochastic approximation of the temporal progression of states which manifest in the original dynamics, with the exception of the slightly jagged and noisy appearance of the trajectories. The topology of the surrogate attractor is enforced by the data reassignment process because all points in $\hat{x}$ occur somewhere in $x$. However, the random walk and the randomness in the data reassignment procedure do not guarantee that the density of an attractor reconstructed from the surrogate time series will be consistent with the original attractor. In the case of both systems, to the extent that the density of the surrogate attractor appears similar to the original, we can infer that the constrained random walk has sampled the attractor in a manner that is relatively consistent with the original dynamics, or alternatively that the random walk symbolic dynamics $\hat{s}$ has a stationary distribution similar to that of the original symbolic dynamics $s$.

Figure 5.3 shows surrogate Lorenz and Rössler time series for various $m$ using constrained random walks on ordinal networks. For the Lorenz time series, as shown in Figure 5.3a, the networks were constructed using lag $\tau = 7$. The surrogate time series for $m = 7$ and $m = 10$ are arguably pseudo-periodic but they do not appear to be useful stochastic approximations of the original system. This is the result of degeneracies which arise when the partition defined by the ordinal map is not sufficiently fine to capture the complexity of the dynamics. For example, if the total time span of a symbol, given $m$ and $\tau$, is shorter in duration than a segment of the Lorenz time series within which the trajectory is cycling around only one of the wings on the attractor, then the ordinal symbolic dynamics $s$ will not distinguish between the two wings. Evidently, this will not only affect both the map from $x$ to $s$ and the reassignment from $\hat{s}$ to $\hat{x}$, but also diminish the quality of the temporal structure encoded in the network and, hence, affect the random walk. Information is lost at each of these steps and the resulting model is less accurate (e.g. both the random walk and the data reassignment procedure can generate trajectories that spuriously jump between the wings of the attractor). By increasing $m$ the ordinal partition becomes finer, and for $m = 13, 16$ and $20$ in Figure 5.3a the surrogate time series
5.3. IN-SAMPLE ANALYSIS OF ORDINAL SURROGATE TIME SERIES

appear to be good stochastic approximations of the deterministic chaotic flow data. Results for the Rössler system, as shown in Figure 5.3b are similar. However, the surrogate time series appear to be a reasonable approximation of the original data for lower values of $m$ than for the Lorenz time series.

![Figure 5.1](image1.png)  
(a) A time delay reconstruction of the Lorenz attractor using the $x$-component time series with embedding dimension $m = 3$ with embedding lag $\tau = 7$. (b) The ordinal network constructed from the same Lorenz time series using embedding dimension $m = 16$ and embedding lag $\tau = 7$. (c) The time delay reconstruction of a stochastic approximation of Lorenz system generated by taking a constrained random walk on the ordinal network and assigning corresponding time series points from the original data.

![Figure 5.2](image2.png)  
(a) A time delay reconstruction of the Rössler attractor using the $x$-component time series with embedding dimension $m = 3$ with embedding lag $\tau = 8$. (b) The ordinal network constructed from the same Rössler time series using embedding dimension $m = 14$ and embedding lag $\tau = 8$. (c) The time delay reconstruction of a stochastic approximation of Rössler system generated by taking a constrained random walk on the ordinal network and assigning corresponding time series points from the original data.

83
CHAPTER 5. GENERATING SURROGATE TIME SERIES FROM ORDINAL NETWORK

Figure 5.3: (a) Surrogate time series $\hat{x}$ of length $L = 5000$ for the $x$-component of the Lorenz system which have been regenerated using constrained random walks on ordinal networks with embedding lag $\tau = 7$ and various embedding dimensions $m$. A segment of the original time series $x$ from which the ordinal network was constructed is shown for reference in the uppermost panel. (b) Surrogate time series $\hat{x}$ of length $L = 2500$ for the $x$-component of the Rössler system which have been regenerated using constrained random walks on ordinal networks with embedding lag $\tau = 8$ and various embedding dimensions $m$. A segment of the original time series $x$ from which the ordinal network was constructed is shown for reference in the uppermost panel.
5.3. Order recurrence plots and recurrence quantification analysis

An order recurrence plot (ORP), as proposed by Groth [2005] and subsequently detailed by Marwan et al. [2007], is defined by the matrix:

$$ R_{n,k} = \begin{cases} 
1 & \text{if } s_n = s_k, \\ 
0 & \text{if } s_n \neq s_k, 
\end{cases} \quad \text{for } n,k = 1,\ldots,L, $$

(5.2)

where $s_n$ and $s_k$ are symbols in the ordinal symbolic dynamics $s$ of length $L$. This is to say that an ORP is equivalent to a traditional recurrence plot, as devised by Eckmann et al. [1987], except that rather than applying a threshold to an appropriate distance metric to define a recurrence between states, the state of the system at time index $n$ is considered to have a recurrence in an ORP at time index $k$ if the states at $n$ and $k$ map to the same ordinal symbol. The primary reason we employ ORPs in this study is not to attempt analysis or classification of time series. Instead, we use them as a visual and quantitative tool for comparing the recurrence properties of the symbolic dynamics $s$ from the original time series $x$ against the symbolic dynamics $\hat{s}$ that are generated by the constrained random walk algorithm.

Figure 5.4 shows the ORPs for $s$ and $\hat{s}$ respectively, computed from the Lorenz time series. We elected to use embedding dimension $m = 16$ based on our qualitative analysis of constrained random walk time series in Section 5.3.1. Both ORPs are characterised by complex structures of short diagonal lines of various lengths which are parallel to the line of identity indicating the possibility of deterministic quasi-periodic dynamics (Marwan et al. [2007]). These structures tend to fall into epochs that are disrupted where the time series spirals outward around one of the unstable foci for at least several cycles. For example, see the regions of low recurrence (white bands) at $n \approx 4000$ and 5000 in the ORP for the original time series and $n \approx 1800$ and 6000 in the ORP from the constrained random walk. In the ORP for the original time series, the epochs appear to be larger and comprise more complex structures than those which occur in the ORP for the random walk. However, there are some small scale structures which are seemingly identical in both ORPs. For example, see the small structure at $(n,k) \approx (5000,5000)$ in the ORP for the original time series and $(n,k) \approx (6000,6000)$ in the ORP from the constrained random walk. In the case of the Rössler time series, the ORPs for $s$ and $\hat{s}$ shown in Figure 5.5 are arguably more similar than those for the Lorenz data. For example, observe the region of low recurrence at $n \approx 4000$ and of high recurrence for $6000 \lesssim n \lesssim 7000$ in the original and surrogate symbolic dynamics. However, close inspection reveals differences in recurrence patterns on short time scales.
Figure 5.4: Order recurrence plots for (a) ordinal symbolic dynamics $\mathbf{s}$ for the chaotic Lorenz time series as computed for $m = 16$ and $\tau = 7$, and (b) a single realisation of a constrained random walk symbolic dynamics $\hat{\mathbf{s}}$ from the corresponding ordinal network model. A recurrence (black) is marked at coordinate $(n, k)$ if the ordinal symbol at time index $n$ recurs at time index $k$ (see Equation 5.2).

Figure 5.5: Order recurrence plots for (a) ordinal symbolic dynamics $\mathbf{s}$ for a chaotic Rössler time series as computed for $m = 14$ and $\tau = 8$, and (b) a single realisation of a constrained random walk symbolic dynamics $\hat{\mathbf{s}}$ from the corresponding ordinal network model.
Recurrence quantification analysis (RQA) measures have been computed on the ORPs for $\mathbf{x}$ and 10 realisations of random walks $\mathbf{\hat{s}}$ for $m = 3, \ldots, 20$, with respect to both the Lorenz and Rössler time series. The results were validated using the Cross Recurrence Plot Toolbox Version 5.21 (R31c) (Marwan et al. [2007]) and are shown in Figures 5.6 and 5.7. The recurrence rate of $\mathbf{\hat{s}}$ is very close to that of $\mathbf{s}$ for all $m$ with negligible variance between the realisations of the random walk for both systems. The two measures of vertical line structures — the laminarity and trapping time — are also in close agreement between the original time series and the constrained random walks.

Turning our attention to the results for the Lorenz time series, the entropy of the distribution of diagonal line lengths are also in close agreement between $\mathbf{s}$ and $\mathbf{\hat{s}}$. Furthermore, determinism and average diagonal line length indicate that diagonal line structures are relatively consistent for smaller $m$. However, these measures suggest significant differences in the dynamics for $m \gtrsim 15$. We postulate that the divergence of $\text{DET}$ and $L$ is likely a result of the ordinal network model breaking down due to the finiteness of data with respect to $m$. As $m$ increases, the number of distinct ordinal symbols encoded from the time series begins to approach the total length of the time series. Therefore the amount of available data no longer guarantees good sampling of the transitional dependencies between symbolic states.

The same discrepancy between the model surrogates and the original data for larger $m$ can be observed in the results for the Rössler time series with respect to the measure $L$. Interestingly, the determinism of the ORP ($\text{DET}$) for the surrogate symbolic dynamics is consistently greater than for the original data, which suggests that the ordinal network is more predictable than the chaotic Rössler system which it models, despite the former being a stochastic process. Finally, it can be observed that the entropy of the distribution of diagonal line lengths for the surrogate symbolic dynamics differs considerably and inconsistently with respect to the original data.

Our results have shown that the RQA measures for recurrence rate and vertical line structures are generally consistent between the original symbolic dynamics and the constrained random walks. Measures of diagonal line structures and the qualitative dissimilarities evident in the ORPs reveal differences in the dynamics between the deterministic chaotic flow generated by the Lorenz and Rössler systems, and the stochastic approximations encoded in their respective ordinal networks. However, the ORP for the constrained random walk is not vastly different to the original and still exhibits structures that would usually indicate quasi-periodic determinism.
Figure 5.6: Recurrence quantification analysis measures plotted against embedding dimension $m$ with $\tau = 7$ for ordinal symbolic dynamics $s$ from the original chaotic Lorenz time series (solid lines), and constrained random walk symbolic dynamics $\hat{s}$ from the corresponding ordinal network model (dashed lines). Each data point for the $\hat{s}$ curves is the sample mean of the measure for 10 realisations of the random walk. Error bars give the standard deviation. The measures computed are the recurrence rate ($RR$), determinism ($DET$), laminarity ($LAM$), average diagonal line length ($L$), trapping time ($TT$), and the Shannon entropy of the distribution of diagonal line lengths ($ENTR$).
Figure 5.7: Recurrence quantification analysis measures plotted against embedding dimension $m$ with $\tau = 8$ for ordinal symbolic dynamics $s$ from chaotic Rössler time series, and constrained random walk symbolic dynamics $\hat{s}$ from the corresponding ordinal network model (as per the ORPs in Figure 5.5). These results have been computed and presented as per those shown in Figure 5.6.
5.3.3 Estimation of Lyapunov exponents

Figure 5.8 shows the logarithm of the average separation of nearby trajectories in the reconstructed phase space for the original time series $x$ and a single surrogate realisation $\hat{x}$ of length $L = 10000$. Note that we have used the algorithm proposed by Rosenstein et al. [1993] (described in Section 1.1) as implemented in the TISEAN software package by Hegger et al. [1999] for the computations in this subsection. The existence of a positive linear scaling region that is invariant to the embedding dimension of the reconstructed phase space is evidence of the exponential divergence of trajectories, where the gradient of this region corresponds to the largest Lyapunov exponent.

A scaling region is clearly observable in Figure 5.8a for $x$ from the Lorenz system, and spans approximately $1 \leq \Delta t \leq 4$. A scaling region of about half this time span can also be observed for $\hat{x}$. The surrogate time series is not deterministic, therefore it would not be correct to infer the identification of chaotic exponential separation of trajectories. However, the presence of a well defined scaling region in $\hat{x}$ with a gradient very similar to that which was computed for $x$ demonstrates that the ordinal network and constrained random walk is partially preserving dynamical properties which relate to invariant measures on the original system.

Apart from the length of the scaling region, the only other difference of note between the curves in Figure 5.8a for $x$ and $\hat{x}$ is the average initial separation distance. This is likely due to an increased degree of heterogeneity in the reconstructed attractor for $\hat{x}$ than for $x$ which results from the random walk. The random walk will be inherently more likely to follow the symbolic trajectories that occur more frequently in the original time series. Therefore, the reconstructed attractor for the surrogate time series will trace the denser areas of the original attractor, but visit the less dense areas with lower probability than would occur in the true dynamics. This is can be observed in Figure 5.1 under close inspection. The topology of the attractor may be such that the Rosenstein algorithm needs to search within a larger distance on average to find a sufficient number of neighbouring states in phase space for the robust estimation of the separation of trajectories and, therefore, the average initial separation of points in the neighbourhood may be larger.

We repeated the estimation of $\lambda_1$ for 10 realisations of $\hat{x}$ of length $L = 10000$ from constrained random walks on ordinal networks that were constructed from the Lorenz time series using parameters $m = 8, 10, 12, 14, 16$ and $\tau = 7$. To reiterate, $\lambda_1$ can only be interpreted as an estimate of the largest Lyapunov exponent for the original time series $x$. For $\hat{x}$ it is simply an estimate of the average rate of divergence nearby trajectories in a given realisation of the time series. We report that no clear linear scaling region can be identified for $m < 12$. Furthermore, we report that the length of the scaling region for $\hat{x}$ is consistently smaller than for $x$ but tends to increase with $m$. Table 5.1 shows that where a linear scaling region could be identified ($m = 12, 14, 16$) the mean estimate of $\lambda_1$ is close to the value estimated for the original time series and variance between the 10 realisations is relatively small. The length of scaling region for these estimates was relatively consistent and generally spanned 1 to 1.5 units. In summary, we find that the ordinal network model of the Lorenz system produces stochastic time series which approximate the exponential separation of trajectories when $m$ is sufficiently large.

The results for the Rössler time series are shown in Figure 5.8b. Observe that it is difficult to identify a single invariant scaling region from the surrogate. We have not attempted calculating and comparing gradients because there are at least two distinct linear scaling regions can be seen for
\( \Delta t \lesssim 10 \) and \( 10 \lesssim \Delta t \lesssim 20 \), and both are no more than half the length of the scaling region for the original time series. What is clear is that the gradient is much lower for the surrogate which implies slower divergence of nearby trajectories and, hence, greater predictability. This result agrees with our observation from Section 5.3.2 that the determinism of the ordinal recurrence plots for ordinal network surrogates of the Rössler time series had consistently higher determinism than the original data. This evidence demonstrates a discrepancy between the original system and the model.

Table 5.1: Estimates of the largest Lyapunov exponent \( \lambda_1 \) as per Figure 5.8a for the original chaotic Lorenz time series \( x \), and 10 realisations of a constrained random walk surrogate time series \( \hat{x} \). The random walk time series \( \hat{x} \) were generated by an ordinal network model of the Lorenz time series using \( m = 12, 14, 16 \) and \( \tau = 7 \). All time series were embedded with \( d = 8 \) and \( \tau = 7 \) for the computation of \( \lambda_1 \).

<table>
<thead>
<tr>
<th>Original time series ( x )</th>
<th>( \lambda_1 = 0.8826 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random walk time series (10 realisations)</td>
<td>( \text{Mean}(\lambda_1) )</td>
</tr>
<tr>
<td>( \hat{x} ) for ( m = 12 )</td>
<td>0.8298</td>
</tr>
<tr>
<td>( \hat{x} ) for ( m = 14 )</td>
<td>0.7250</td>
</tr>
<tr>
<td>( \hat{x} ) for ( m = 16 )</td>
<td>0.9784</td>
</tr>
</tbody>
</table>

Figure 5.8: (a) The average separation of nearby trajectories \( \bar{q} \) against \( \Delta t \) in the reconstructed attractors of the original chaotic Lorenz time series \( x \), and a single realisation of a constrained random walk surrogate time series \( \hat{x} \) for the estimation of the largest Lyapunov exponent \( \lambda_1 \). The random walk time series \( \hat{x} \) was generated by an ordinal network model of the Lorenz time series using \( m = 16 \) and \( \tau = 7 \). For both time series, \( \bar{q} \) is plotted for \( \tau = 7 \) and \( d = 3, 4, 5, 6, 7, 8 \) (from lowest to highest curve in each set respectively) where \( d \) is the dimension in which \( x \) and \( \hat{x} \) are embedded for the computation of the separation of trajectories. The minimum spatial scale on which the algorithm searches for neighbours is given by \( \epsilon_{\text{min}} \). The estimates for \( \lambda_1 \) are given by the gradients of the dashed lines where the lower line corresponds to \( x \) and the upper line to \( \hat{x} \). These lines are computed using linear regression based on the points in the linear scaling regions of \( \bar{q} \) which are demarcated by the solid portions of each respective line of best fit. (b) Results for the same numerical analysis preformed for the chaotic Rössler time series \( x \). The random walk time series \( \hat{x} \) was generated by an ordinal network model of the Rössler time series using \( m = 14 \) and \( \tau = 8 \). Linear regression analysis is not shown in this case due to the significant difference between the original data and the surrogate.
5.4 Out-of-sample prediction using ordinal networks

Prediction is a classic problem in nonlinear time series analysis. Having established in the previous section that ordinal network models preserve the in-sample properties of nonlinear time series to some extent, we now present a brief investigation of the out-of-sample predictive properties of ordinal network models for low dimensional chaotic systems.

5.4.1 Prediction algorithm

Herein we define an algorithm for predicting the future state of a system from the univariate time series \( x = \{x_n\}_{n=1}^N \) using an ordinal network model. The objective is to compute an estimate of the expected value of an ensemble of random walk surrogate time series \( \hat{x} = \{x_{n+L}^N\}_{n=N+1} \) for \( L \geq 1 \) using a Markov chain Monte Carlo approach, where the random walk is initialised such that it always begins from the final state observed in the data. Note that we maintain consistent time indices between the end of the data and the beginning of the prediction.

We begin the predictive algorithm by generating the ordinal symbolic dynamics and \( s = \{s_n\}_{n=1}^{N-(m-1)\tau} \) and the ordinal network given by adjacency matrix \( A \) for the embedding parameters \( m \) and \( \tau \). The random walk on \( A \), given by \( \hat{s} = \{s_n\}_{n=N-m\tau+1}^{N+L} \), is initialised with the node corresponding to the final \( \tau \) symbols in \( s \), explicitly: \( \{s_n\}_{n=N-m\tau+1}^{N-(m-1)\tau} = \{s_n\}_{n=N-m\tau+1}^{N-(m-1)\tau} \). It should be apparent that while the random walk starts at time index \( N-(m-1)\tau + 1 \), the random walk is only predictive for \( n > N \). We then proceed with the random walk as per the algorithm defined in Section 5.2.2 from line 4 onwards with minor modifications to the algorithm.

These modifications are, firstly, that the time indices of the walk are shifted such that the walk begins at \( PTR = N-(m-1)\tau \) and continues until the walk reaches the desired length \( PTR = N+L \) or until all possible paths are exhausted and a walk of the desired length \( L \) is not possible. The latter case can be detected if \( PTR < N-(m-1)\tau \), at which point the algorithm breaks. Secondly, recall that in Section 5.3 we were only interested in generating surrogates. Therefore, where there did not exist any possible random walk path of the desired length from the specified initial conditions, it was sufficient for our purposes to randomly select a new set of initial conditions from the original symbolic dynamics and retry the random walk until it became possible to generate a random walk of sufficient length. In this section, our aim is to make a prediction, therefore, if there does not exist a possible random walk up to the desired prediction time index \( N+L \) from the initial conditions given by the final \( \tau \) symbols in \( s \), then we step our initial condition back along the symbolic dynamics by a single time index and reattempt the random walk. If a random walk of the desired length is still impossible then we continue to step the initial condition backwards along the symbolic dynamics until a random walk of the desired length is possible. This will almost certainly increase the expected error in the prediction because the random walk has to evolve further in time before we reach the domain of time indices which constitute a prediction.

Once we attain the random walk \( \hat{s} \) we reconstruct a time series using the data reassignment procedure detailed in Section 5.2.3. We then repeat this process \( k \)-times to generate an ensemble of random walk predictions. Finally, we define the ordinal network model prediction of the time series to be the expected value (mean) of the ensemble of random walk predictions \( E[\hat{x}] \).
5.4.2 Numerical investigation

To investigate the performance of the ordinal network predictive algorithm, we generate time series data by numerically solving the Lorenz system (equation 5.1) with parameters as per Section 5.3. The total length of the time series is 101000 observations which are discretely sampled with a time step of $\Delta t = 0.01$ which corresponds to approximately 100 points per mean cycles. This is divided into training data of length $N = 100000$ and out-of-sample test data of length $L = 1000$. We then construct an ordinal network model from the test data with embedding parameters $m = 16$ and $\tau = 7$ and compute an ensemble of $k = 50$ random walk predictions.

Figure 5.9 shows a representative example of an ordinal network model out-of-sample prediction for a $x$-component time series from the Lorenz system. Observe that the prediction remains close to the true trajectory for several cycles before beginning to diverge as it traverses the separatrix. The predicted trajectory then approaches to the mean value of the time series data. This occurs due to the mixing of the Markov process because in the limit of $n \to \infty$, and for a large ensemble of random walks, the distribution of symbolic states over the ensemble at a given time step should converge to the stationary distribution of the Markov chain. Given that the data reassignment procedure reconstructs the prediction $E[\hat{x}]$ using embedded state vectors selected at random from the training data $x$ from the set corresponding to each symbol (e.g. contained within each ordinal partition element), the distribution of scalar values in $E[\hat{x}]$ should be close to that of $x$.

![Figure 5.9: Out-of-sample prediction for an $x$-component time series $x$ from the Lorenz system using surrogates generated by a constrained random walk on an ordinal network. The ordinal network model was constructed from training data using an embedding dimension of $m = 16$ and embedding lag $\tau = 7$. The black curve is the out-of-sample data from $x$ against which the model prediction is compared. The dashed yellow curve is a single realisation of a surrogate $\hat{x}$ from a random walk on the network that was initialised with the final $\tau - 1$ length in the symbolic dynamics $s$ enumerated from the training data. This specific realisation happens to be highly correlated with the out-of-sample data but in itself should be considered a simulation, not a prediction. The set of grey curves make up an ensemble of 50 realisations of surrogate time series generated from this same initial condition. The purple curve shows the ordinal network model prediction which is an estimate of the expected value of the surrogate $\hat{x}$ computed as the mean value over the ensemble of realisations.](image-url)
To quantify the expected model error we generate 100 realisations of the Lorenz time series, compute the ordinal network model prediction using \( m = 16 \) and \( \tau = 7 \), and then compute the mean of the linear cross-correlation and root mean squared error (RMSE) for all predictions with respect to the true trajectory. It is necessary to consider multiple realisations of the chaotic Lorenz data because in the context of this numerical investigation the model error can arise from any combination of three primary sources: finiteness of data which results in an incomplete description of the dynamics; heterogeneity of the attractor and the local dynamical properties of the region of the attractor where prediction begins; or the capacity of the ordinal network to accurately model the dynamics of the time series. To provide a benchmark for the performance of the ordinal network prediction, and given that we expect this prediction to converge on the mean of the training data \( \mathbf{x} \), we also compute the RMSE error between the true trajectory and mean of \( \mathbf{x} \), which we call the mean predictor.

Results are shown in Figures 5.10 and 5.11 for cross-correlation and RMSE respectively. The mean cross-correlation is high \( (r(\mathbf{x}, \mathbf{E}[\hat{\mathbf{x}}]) > 0.9) \) and mean RMSE relatively low for predictions of \( \Delta n < 100 \) steps. However, observing the results for the complete ensemble of data/model realisations shows that in very rare cases the model performs very poorly with \( r(\mathbf{x}, \mathbf{E}[\hat{\mathbf{x}}]) < 0.8 \) and RMSE larger than for the mean predictor for predictions less than \( \Delta n < 50 \) steps. Therefore, while the ordinal network model clearly has the capacity to produce a reasonable short term prediction for most realisations of a low dimensional chaotic system, this approach should be used cautiously if implemented. We reiterate here that the purpose of this investigation was simply to test the extent to which the ordinal network model can be used for out-of-sample prediction. It is unlikely that a stochastic model such as this would provide a better prediction for a low dimensional chaotic system than the raft of existing methods which have been specifically developed to model and predict nonlinear dynamics from time series (e.g. local or global nonlinear fits, radial basis functions etc.).

5.5 Chapter summary

In this chapter we have constructed ordinal networks using discrete sampled time series from continuous chaotic Lorenz and Rössler systems, and then used random walks on the network models to regenerate surrogate time series data and study the model’s out-of-sample predictive properties. Our random walk algorithm is not a first order Markov chain because it requires a history of past symbolic states. Symbolic states spaced by the first \( 1, ..., (m - 1) \) multiples of \( \tau \) will contain overlapping elements. The algorithm has been designed to constrain the walk such that the overlapping elements in these ordinal patterns have congruent ordering so that random walk surrogates will adhere more closely to the dynamics of the original time series. Our results show that ordinal network models can regenerate time series which both approximate and predict the dynamics of the original time series for a sufficiently large embedding dimension \( m \).

In-sample analysis of surrogates revealed that the order recurrence plots for both original and random walk symbolic dynamics comprise similar structural patterns which were predominately characterised by short diagonal lines of various lengths as expected given the quasi-periodic dynamics. However the order recurrence plots for constrained random walks appear to exhibit more regularity. Furthermore, while the recurrence quantification analysis measures of vertical line structures are relatively consistent between the original and random walks, measures of diagonal line structures
5.5. CHAPTER SUMMARY

**Figure 5.10:** Cross correlation between the ordinal network model prediction and the out-of-sample test data for 100 different realisations of the training/test data $\mathbf{x}$ from the Lorenz system. The ordinal network model parameters and procedure of computing the model prediction are both as per Figure 5.9. Each grey curve shows the Pearson correlation coefficient $r(\mathbf{x}, \mathbf{E}[\hat{\mathbf{x}}])$ between the model prediction $\mathbf{E}[\hat{\mathbf{x}}]$ and the out-of-sample test data from a single realisation of the training/test data $\mathbf{x}$ for a predicted trajectory of length $\Delta n$. The blue curve gives the mean of the correlation between the model prediction and the test data over the ensemble of realisations.

**Figure 5.11:** Root mean squared error (RMSE) between the ordinal network model prediction and the out-of-sample test data for 100 different realisations of the training/test data $\mathbf{x}$ from the Lorenz system. The ordinal network model parameters and procedure of computing the model prediction are both as per Figure 5.9. Each grey curve shows the RMSE between the model prediction and the out-of-sample test data for a single realisation of the training/test data $\mathbf{x}$ for a $\Delta n$ step prediction. The green curve gives the mean of the RMSE between the model prediction and the test data over the ensemble of data/model realisations. For reference, the dashed black curve shows the RMSE between the a mean-predictor (the mean value of the training data) and the out-of-sample test data over the ensemble of data/model realisations.
diverge as the size of the ordinal network approaches the length of available data. We also investigated the average separation of nearby trajectories in network generated surrogate time series for constrained random walks. Our results indicate that the dynamics of the surrogates approximate the deterministic exponential divergence characteristic of the original time series, hence, the network model preserves some information pertaining to the largest Lyapunov exponent. However, the observable scaling region for the surrogates is consistently shorter than for the original time series.

We also proposed a method for out-of-sample prediction using constrained random walks and Markov chain Monte Carlo analysis to estimate the expected value of a scalar time series constructed from the ordinal network model. Numerical investigations showed that the ordinal network model was a significantly better predictor than the mean of the training data for low dimensional continuous chaotic time series generated by the Lorenz system. Predictions were generally accurate while trajectories remained in one wing of the attractor but diverged rapidly over several cycles when the true trajectory passed near or through the separatrix.
Chapter 6

Counting forbidden patterns in irregularly sampled time series

ABSTRACT

Experienced practitioners of nonlinear time series analysis will understand the importance of discerning deterministic dynamics from stochastic processes using measured data. The reasons are, firstly, that many of the existing techniques in the field are theoretically dependent on the assumption of determinism. This includes any method requiring a phase space reconstruction using time delay embedding such as the estimation of correlation dimension or Lyapunov exponents. Secondly, knowing from experimental data that a system is likely driven by a deterministic process can be used to justify the construction of deterministic models. It has been established in the literature that the count of ordinal symbols which do not occur in a time series, called forbidden patterns, is an effective measure for the detection of determinism in noisy data. A very recent study by Kulp et al. [2016] has shown that this measure is also partially robust against the effects of irregular sampling. In this chapter we extend said research with an emphasis on exploring the parameter space and find that the measure is more robust to under-sampling and irregular sampling than previously reported. Using numerically generated data from the Lorenz system and the hyper-chaotic Rössler system, we investigate the reliability of the relative proportion of ordinal symbols in periodic and chaotic time series for various degrees of under-sampling, random depletion of data, and timing jitter. Discussion and interpretation of results focuses on determining the limitations of the measure with respect to optimal parameter selection, the quantity of data available, the sampling period, and the Lyapunov and de-correlation times of the system.

This chapter is structured as follows. Section 6.2 provides a brief review of the concept of forbidden patterns as a means for detecting determinism. Section 6.3 comprises the description of methodology, results and discussions for three separate numerical investigations into the effect of sampling on the count of forbidden patterns as performed on time series from the Lorenz system and the hyper-chaotic Rössler system, namely: under-sampling (6.3.1), random depletion of data (6.3.2), and timing jitter (6.3.3).
CHAPTER 6. COUNTING FORBIDDEN PATTERNS IN IRREGULARLY SAMPLED TIME SERIES

6.1 Introduction

Irregularly sampled time series are commonplace in data that has been measured beyond the confines of controlled laboratory experiments. For example, time series can be affected by timing jitter in measurement devices which can induce bias in amplitude measurements of sampled waveforms (Souders et al. [1990]). More extreme cases can arise where measurements can only be taken in particular weather conditions or when some remote sensing apparatus is within range of communication such as astrophysical data (Scargle [1982]) which must be recorded with telescopes or received from satellites. There are also types of data which are inherently irregularly sampled such as financial time series where stock prices are updated when transactions are made (Suzuki et al. [2010]), or recordings of biological signals such as cardiac inter-beat intervals (Berntson and Stowell [1998]; Xia et al. [1992]).

The reality of irregular sampling compels us to investigate the reliability of any measure or method that we would apply to analyse such data. In a recent paper by Kulp et al. [2016], the authors undertook a study of the robustness of the count of forbidden patterns with respect to several irregular sampling schemes. Recall from Chapter 3 that a forbidden pattern is any ordinal symbol in the set of permutations of length $m$ which does not occur in the ordinal symbolic dynamics $S$ corresponding to a given time series $x = \{x_n\}_{n=1}^{N}$, and that confirming the existence of forbidden patterns for a given time series is a theoretically sound and practically applicable criterion for asserting that the data contains a deterministic component (Amigó et al. [2007, 2010]; Rosso et al. [2012]). Kulp et al. [2016] used numerically generated time series from periodic and chaotic regimes of the Lorenz system to test the effects of irregular sampling on the count of forbidden patterns and found that the measure was robust for the sampling schemes considered in their study given a limited degree of irregularity, but that more extreme fluctuations in the sampling period would cause misleading results. However, they elected to use a very narrow parameter range for the computation of ordinal patterns with respect to the pattern length and focussed primarily on qualitative discussion of the frequency content of the signals and the Nyquist frequency obtained from Fourier transforms of the time series to explain their results.

Our investigation is intended to serve as an extension to Kulp et al. [2016] by virtue of several key distinctions. Firstly, we study the benefits and potential limitations of exploring a larger parameter space for the ordinal pattern length. Secondly, and critically, we present the interpretation of our results and formulate our conclusions through the lens of nonlinear dynamical systems theory, which is not only novel and complementary, but also a necessary perspective in light of the much broader motivation: to detect quantitative evidence of deterministic dynamics in complex systems based on measured time series data. Differences in our approach that are of a technical nature will be detailed throughout this chapter within the context in which they arise. The scope of this chapter is limited to studying the count of forbidden patterns for data which is under-sampled, randomly depleted or affected by timing jitter respectively. Refer to Figure 6.1 for an example schematic of the irregular sampling processes.

6.2 Counting forbidden patterns

Consider an arbitrary time series $x = \{x_n\}_{n=1}^{N}$ that has been discretely sampled from a continuous system. We begin by enumerating the symbolic alphabet $S$ as described in Section 2.1 which is the
6.2. COUNTING FORBIDDEN PATTERNS

Figure 6.1: An example schematic depicting possible realisations of irregular sampling processes. Data is randomly depleted by the removal of a random percentage of observations from the time series. The rate of depletion in this example is 30%. Timing jitter is simulated by perturbing a set of uniformly spaced sampling times by realisations of a random variable drawn from the normal distribution $N \left(0, \sigma^2\right)$. The variance of the jitter in this example is $\sigma^2 = 0.2(\Delta t)$.

unique set of ordinal patterns in the ordinal symbolic dynamics $s$. In this chapter we consider the embedding dimension $m$ as a free parameter and fix the embedding lag $\tau = 1$. Henceforth we use the terms ordinal pattern and ordinal symbol interchangeably.

Recall that for a given window length there are a total of $m!$ unique ordinal patterns that can possibly occur in our time series. It is intuitive that all $m!$ patterns will almost certainly occur in a time series generated by a stochastic process for $N \to \infty$. Perhaps less intuitive but pivotal in the context of this Chapter is that for time series generated by deterministic dynamics there will exist a set of forbidden patterns which can never occur (Amigó et al. [2014]). Using analytical arguments based on the concept of topological permutation entropy, Amigó et al. [2010] demonstrated that forbidden patterns will always exist for deterministic one dimensional piecewise monotone interval maps, and numerical investigations by Kulp et al. [2016] suggest that forbidden patterns are also characteristic of continuous chaotic dynamics. By virtue of this phenomenon it is possible to use the count of the forbidden patterns to detect determinism in time series data. Complications arise from the fact that real data is finite, hence an ordinal pattern that is admissible in the dynamics may not occur during the period of observation. However, methods have been developed in with this issue in mind by Amigó et al. [2007, 2008, 2010] which are effective for the detection of determinism in relatively short noisy time series.

Having discussed the application of forbidden patterns for the detection of determinism in noisy time series, we return to the central question of this study: how reliable is the count of forbidden patterns from time series when the data is under-sampled or irregularly sampled? For the remainder of this study we will refer to the measure:

$$P_f^{(m)} = \frac{m! - |S|}{m!}.$$
which is the relative number of forbidden patterns with respect to the total number of possible patterns dependent on \( m \), where \(|.|\) denotes cardinality. This measure is an estimator of the relative number of true forbidden patterns. Alternatively, \( 1 - P_f^{(m)} \) is the proportion of patterns admissible by the dynamics which would almost certainly occur as \( N \to \infty \). In the following investigation we test the robustness of \( P_f^{(m)} \) with respect to various sampling schemes for a range of \( m \).

### 6.3 Numerical investigation

Before proceeding with the main line of inquiry we present Figure 6.2 as a result which motivates the study of \( P_f^{(m)} \) for larger \( m \). Kulp et al. [2016] limited the scope of their investigations to \( m \leq 5 \). However, it is clear from Figure 6.2 that \( P_f^{(m)} \) also appears to be a more robust measure against both noise and false forbidden patterns as \( m \) increases (provided \( N \gg m! \)). This is due to the outgrowth property, first reported by Amigó et al. [2007], which causes forbidden patterns to dominate (super-exponential growth) with respect to admissible patterns (exponential growth) as \( m \) increases for deterministic time series. Moreover, there is a clear difference in the curve for the chaotic Rössler system and the respective curves for the chaotic Lorenz and Chua systems. While smaller \( m \) appears to be sufficient for detecting determinism in periodic and chaotic time series generated by the Rössler system, time series from the Lorenz and Chua circuit attractors require \( m \geq 5 \) before the measure would be a reliable estimator in the presence of even small amounts of noise. This is almost certainly due to the more complex two scroll attractors in the Lorenz and Chua systems which are generating more unique ordinal patterns from their respective trajectories for a given \( m \) than the single scroll phase coherent Rössler attractor.

![Figure 6.2: Percentage of forbidden patterns against pattern order \( m \) in long \((N \gg m!)\) highly sampled time series from periodic and chaotic regimes of archetypal chaotic systems.](image)

The method of investigation used for this chapter generally follows the approach taken by Kulp et al. [2016] but with several fundamental exceptions. Firstly, we have elected to report the relative number of forbidden patterns as a percentage rather than an absolute count, thereby enabling easier comparison between results for different \( m \). Secondly, Kulp et al. [2016] specified a fixed time series length \( N \) while varying the sampling period \( \Delta t \). This meant that the total integration time length...
6.3. NUMERICAL INVESTIGATION

of each time series was directly proportional to the experimental variable $\Delta t$ and changed with each simulation. Instead, we opt to fix the integration time length and allow $N$ to vary. The practical analogy for our approach, distinct from the former, is that we are constrained to observe a system for some fixed time, hence the system generates some fixed total amount of information. We then concern ourselves with the question of how highly or regularly the system should be sampled to extract information of significant quantity and quality for $P_f^m$ to be a reliable estimator.

We perform our investigation on highly sampled time series generated by the Lorenz system given in Equation 5.1 with parameters $\rho = 200$ and $\beta = 8/3$. To generate periodic time series we set the final parameter $\sigma = 100$. To generate chaotic time series we use $\sigma = 25$. These parameters are selected specifically to match those used by Kulp et al. [2016]. Results are also computed for time series from the hyper-chaotic Rössler system. Time series were generated by solving the systems using a fourth-fifth order Runge-Kutta algorithm and transients were removed.

6.3.1 Under-sampling

To test the effect of under-sampling for a regular sampling period we generate time series for a total integration time length of 20000 time units (approximately $6 \times 10^5$ cycles) for $0.01 \leq \Delta t \leq 0.3$ then compute $P_f^m$ for $4 \leq m \leq 8$. In the periodic case (Figure 6.3a) it can be observed that $P_f^m \gg 0$ as expected, and is stable for $6 \leq m \leq 8$. However, the value of the measure appears to be sensitive to $\Delta t$ in a highly non-trivial manner when $m \leq 5$. This sensitivity is very likely due to aliasing effects as postulated by Kulp et al. [2016], who also reported that when $\Delta t$ is such that the sampling frequency $1/\Delta t$ is less than the Nyquist frequency, results should be discarded because the data is no longer sufficient to describe the waveform of the underlying dynamics. On the latter point, we assert that that $P_f^m$ will be non-zero for all $\Delta t$ given regular sampling period and for sufficient $m$. This is because aliasing will result in a time series with a period greater than the original signal but one which is still strictly periodic, hence forbidden patterns will still exist and be a characteristic indicator for the detection of determinism even as $\Delta t$ becomes large.
CHAPTER 6. COUNTING FORBIDDEN PATTERNS IN IRREGULARLY SAMPLED TIME SERIES

For chaotic time series, results shown in Figure 6.3b show \( P_f^{(m)} \gg 0 \) converging to zero for \( m = 4, \ldots, 7 \). This is best explained by the fact that time series points in each window become decorrelated with increasing \( \Delta t \) due to sensitivity to initial conditions imposed by the chaotic dynamics, and subsequently the time series will resemble a random process when \( \Delta t \) is sufficiently large. We can find an estimate for this value of \( \Delta t \) by computing the autocorrelation function and Lyapunov time of the time series, shown in Figure 6.4. Therefore it should be expected that \( P_f^{(m)} > 0 \) for \( \Delta t \lesssim 0.28 \) and \( P_f^{(m)} \approx 0 \) for any larger \( \Delta t \). Referring back to Figure 6.3b, this suggests that selecting \( m = 7 \) provides the most robust \( P_f^{(m)} \) for the range of \( \Delta t \) where the time series retains evidence of the underlying determinism.

![Figure 6.4: The autocorrelation function and Lyapunov time (vertical dashed line) for the regularly sampled chaotic Lorenz time series. The Lyapunov time of the system was estimated from the time series using the TISEAN Hegger et al. [1999] software package function lyap_k.](image)

However, this also implies an anomaly in the results for \( m = 8 \) because \( P_f^{(m)} \gg 0 \) and does not appear to be converging anywhere close to zero. We have included this curve to highlight the issue of false forbidden patterns and insufficient data. It is generally accepted that ordinal pattern statistics will be reliable if the length of data available meets the heuristic condition \( N \gg m! + m - 1 \) to ensure that it will be likely that most of the admissible patterns occur in the data (Amigó et al. [2007]). In the case of the results for \( m = 8 \) shown in Figure 6.3b, we have \( N \approx 70000 \) for \( \Delta t = 0.28 \) and \( m! = 8! = 40320 \), so this condition is not met as \( \Delta t \) approaches the Lyapunov time. Therefore, it is probable that the estimate of \( P_f^{(m)} \) for \( m = 8 \) includes many false forbidden patterns and, as such, is inaccurate. To confirm this, we compute \( P_f^{(m)} \) with \( m = 8 \) against \( N \leq 10^6 \) for several \( \Delta t \) either side of the Lyapunov time, as shown in Figure 6.5. These results show that \( P_f^{(m)} \) converges to zero (or very close to zero) with increasing \( N \) for \( \Delta t > 0.28 \) and is non-zero for \( \Delta t \) in the lower range where the time series exhibits a reasonable degree of correlation due to determinism. For the remainder of this study we use \( \Delta t = 0.05 \) and 0.1 as our fundamental sampling periods (before depletion or jitter are applied to the data) because \( P_f^{(m)} \) is non-zero and stable up to \( m = 8 \) for \( N > 10^5 \).

Our key findings in this subsection follow. Firstly, detection of determinism using only \( P_f^{(m)} \) will likely fail for chaotic time series when the sampling period exceeds the Lyapunov time or the point at which the autocorrelation function approaches zero. We note here, however, that this finding does
not exclude the possibility that other more rigorous quantitative methods for detecting determinism using forbidden patterns, such as those proposed in Amigó et al. [2008, 2010], may still be effective for sparse irregularly sampled time series. Secondly, selecting larger $m$ will provide better detection of determinism when using forbidden patterns methods until the point where there is insufficient data to ensure accurate sampling of admissible ordinal patterns.

In order to verify our findings with other example systems, we repeated our numerical experiments using $x$-component time series generated by: the Rössler system (equation 2.4) with parameters $\alpha = 0.3848$, $\beta = 2$ and $\gamma = 4$ such that the system is in a period-8 regime; and and the four-dimensional (4D) Rössler system given by the equations:

\[
\begin{align*}
\frac{dx}{dt} &= -y - z, \\
\frac{dy}{dt} &= x + ay + w, \\
\frac{dz}{dt} &= b + xz, \\
\frac{dz}{dt} &= -cz + dw,
\end{align*}
\]

(6.2)

with parameters $a = 0.25$, $b = 3$, $c = 0.5$ and $d = 0.05$ such that the system is in a hyper-chaotic regime (Rössler [1979]). The range of $\Delta t$ values for simulations with these data are selected to ensure an approximate equivalence of sampling intervals with respect to the average number of points per cycle between the different time series data used in this investigation. We do this because one of the central aspects of these investigations is the relationship between the sampling frequency and the time scale of dynamical evolution, as highlighted in Kulp et al. [2016]. Figure 6.6 shows the sensitivity of the relative count of forbidden patterns to the sampling period for the hyper-chaotic 4D Rössler time series. The results here are different from those for the Lorenz time series, in that $P_f^{(m)}$ remains non-zero for large $\Delta t$ where $m \geq 5$. Even in the case of $m = 4$ where the measure does reach zero, it then trends upwards to a non-zero value as $\Delta t$ increases. We postulate that the difference between the results for the two systems arises from their respective autocorrelation functions. As shown in Figure 6.4, the Lorenz time
series decorrelates rapidly to values near zero. On the other hand, the autocorrelation function for the hyper-chaotic Rössler system is highly cyclic with slow decay. This is a reflection of the strongly quasi-periodic motion of the phase-space trajectory with respect to the one-dimensional projection onto the $x$-coordinate. Because ordinal patterns are robust to small changes in amplitude and the system has relatively consistent phase dynamics, the time series retains an element of its characteristic determinism to which ordinal patterns are sensitive, even for very sparse sampling, and hence why $P_f^{(m)}$ is non-zero for large $\Delta t$. However, it is almost certain that $P_f^{(m)}$ would eventually decrease. The dips in $P_f^{(m)}$ for $m \leq 7$ are likely an aliasing phenomena as also observed in the results for the periodic Lorenz time series (Figure 6.3a).

### 6.3.2 Random depletion

For this test we generate time series for a total integration time length of 20000 time units for $\Delta t \leq 0.05$ and then randomly remove data points from the time series. Once more we have followed the methodology of Kulp et al. [2016] but extend the computations to larger $m$.

Figure 6.7 shows $P_f^{(m)}$ with respect to the percentage depleted ($d\%$). Beginning at the leftmost point of the figure, one observes that $P_f^{(m)}$ decreases with $d\%$ for both periodic and chaotic dynamics. The count of forbidden patterns decreases because random depletion introduces a stochastic component in the dynamics by way of random distortion of the temporal correlation between elements in each $m$-length window of successive data points from which the ordinal patterns are computed. This eventually results in time series with no forbidden patterns, as can be observed in the curves for $m = 5, ..., 7$, at which point $P_f^{(m)}$ is no longer a reliable criterion for determinism.

The first feature of these results which is new in this study is that all of the curves ultimately trend upward to $P_f^{(m)} = 100\%$, most notably for $m = 8$ where the forbidden pattern count trends upward at $d\% \approx 80\%$ before it is able to reach zero. This is due to false forbidden patterns becoming prevalent such that the condition for reliable sampling of ordinal patterns $N(1 - d\%) \gg m!$ is no longer met. For example, for $d\% = 90\%$ we have $0.1N = 40000 \approx 8!$ and therefore $d\%$ must be far less than this level for $P_f^{(m)}$ to be a usable measure when $m = 8$. 

![Figure 6.6: Percentage of forbidden patterns against sampling period $\Delta t$ for regularly sampled time series generated by the hyper-chaotic 4D Rössler system.](image)
Figure 6.7: Percentage of forbidden patterns against the percentage of the observations that have been randomly removed from the data for time series generated by the Lorenz system in the (a) periodic and (b) chaotic regimes. The data was sampled with at regular period $\Delta t = 0.05$ prior to the depletion process. Each datum is the mean of 100 realisations of the simulation. Variance across all simulations is less than 2.4%.

Figure 6.8: Percentage of forbidden patterns against depletion percentage for time series generated by the (a) 8-periodic Rössler and (b) hyper-chaotic 4D Rössler time series. The data was sampled at regular periods (a) $\Delta t = 0.31$ and (b) $\Delta t = 0.285$, respectively, prior to the depletion process. Each datum is the mean of 100 realisations of the simulation. Variance across all simulations is less than 0.2%. 

105
The second new finding in this study pertaining to random depletion is that, as per the undersampling test, the results for both periodic and chaotic time series show that $P_f^{(m)}$ is higher for larger $m$ and therefore provides a more robust measure for the detection of determinism. This can be explained by examining the worst case for $P_f^{(m)}$ when the time series is affected by the process of random depletion. For simplicity, assume very long data and a level of depletion such that $N(1 - d\%) \gg (m!)$. Then assume the case that for all $m$-length windows of the data, removing a single data point from within a window will result in a new ordinal pattern that was forbidden before the removal of that data point, and that the pattern originally corresponding to this window is also counted somewhere else in the data. Now consider $Nd\%$ points depleted from the data each spaced by a minimum of $m$ data points such that the maximal number of new patterns, equal to $mNd\%$, is observed. Therefore, forbidden patterns can only be observed if $m[(m-1)! - Nd\%] > 0$. This condition is not intended to serve as a heuristic for electing to use forbidden patterns as a test for determinism because in that capacity it is excessively conservative. However, it implies that for a given length of data and percentage of depletion, using larger $m$ should provide a more reliable count of forbidden patterns where sufficient data is available.

Figure 6.8 shows the corresponding results for the 8-periodic Rössler and the hyper-chaotic 4D Rössler experiments. We select $\Delta t$ such that we sample approximately 20 points per cycle in both cases. Evidently, the results are consistent with our previous analysis for the Lorenz data. We have elected to include the plot for the 8-periodic regime to highlight two observations. Firstly, the large $P_f^{(m)}$ count which characterises periodic systems. Secondly, the robustness of the count even at significant depletion percentages, as illustrated by the very low rate of decrease for $d\% \leq 40\%$ in Figure 6.8a.

### 6.3.3 Timing jitter

In this final test we investigate $P_f^{(m)}$ for time series with simulated timing jitter applied to the sampling period. This is achieved by defining a vector of sampling times with regular period and perturbing each sample time by realisations of a random variable with a normal distribution of zero mean. The intensity of the jitter is controlled by the standard deviation $\sigma$. If a sampling time index is perturbed to the extent that it overlaps with adjacent indices then the vector is sorted to recover temporal ordering. The Runge-Kutta solver returns observations for the periodic and chaotic Lorenz systems at the times specified by the sampling vector for total integration time length of 20000 time units. Therefore, $N = 400000$ and $N \gg m!$ holds for these tests.

The results for the periodic case are shown in Figure 6.9. It can be observed that $P_f^{(m)}$ appears to converge to a steady value when $\sigma \approx \Delta t$. The increasing jitter perturbs the temporal correlation between adjacent data points resulting in more admissible patterns until $\sigma > \Delta t$ when the sampling scheme is essentially random with constant average density on the time scale $\Delta t$. This is likely to be an unrealistic level of jitter for most applications, nonetheless, in this scenario the primary factor affecting the count of forbidden patterns is the average density of sampling. As can be seen by comparing Figure 6.9a with Figure 6.9b, larger sampling periods make $P_f^{(m)}$ less reliable in the presence of timing jitter. However, the key finding is that $P_f^{(m)}$ maintains non-zero values for $\sigma$ as large as $2\Delta t$ and beyond when $m$ is sufficiently large. Kulp et al. [2016] only provides results for the jitter timing test for $m = 5$. 

106
6.3. NUMERICAL INVESTIGATION

Figure 6.9: Percentage of forbidden patterns against the timing jitter for time series generated by a periodic Lorenz system. In this experiment the underlying regular sampling period was set to (a) $\Delta t = 0.05$ and (b) $\Delta t = 0.1$, and the degree to which the regular sampling period was perturbed was governed by $\sim \mathcal{N}(0, \sigma^2)$.

Figure 6.10: Percentage of forbidden patterns against the timing jitter for time series generated by a chaotic Lorenz system. In this experiment the underlying regular sampling interval was set to (a) $\Delta t = 0.05$ and (b) $\Delta t = 0.1$, and the degree to which the regular sampling period was perturbed was governed by $\sim \mathcal{N}(0, \sigma^2)$.
For this choice of $m$, $\mathcal{P}^{(m)}_j$ is only non-zero in the periodic time series for jitter with $\sigma < 0.5$ when $\Delta t = 0.05$.

Results are similar in the chaotic case as observable in Figure 6.10. The estimator $\mathcal{P}^{(m)}_j$ remains a reliable criterion for determinism for sufficiently large $m$. On the other hand, for larger $\Delta t$ the count of forbidden patterns is severely diminished in the presence even for small $\sigma$ (Figure 6.10b), despite this sampling period providing reliable results for regular sampling. This is because the correlation between adjacent data points in a chaotic time series disintegrates rapidly as $\Delta t$ increases due to exponential divergence of trajectories as discussed in Section 6.3.1. Under these dynamical conditions even small perturbations to temporal correlation can be significant. Finally, as was clear for periodic time series, it is reasonable to postulate from these curves that the density of sampling, rather than the regularity of sampling period, should be the primary concern when assessing the likelihood of obtaining a reliable $\mathcal{P}^{(m)}_j$ from experimental data affected by timing jitter.

Figure 6.11 shows the results for the hyper-chaotic 4D Rössler experiment. All key findings are consistent with those for the Lorenz time series. We observe saturation to a limiting value once $\sigma \simeq \Delta t = 0.285$ for $m \geq 7$ and non-zero values for $\sigma > 2\Delta t$ for values of $m$ considered. The estimator $\mathcal{P}^{(m)}_j$ retains its reliability in detecting determinism and is more sensitive to sampling density than the regularity of the sampling period. Note that the sampling frequency of the underlying (unperturbed) regular grid captures approximately 20 points per cycle, which is sufficiently high as shown in Figure 6.6.

### 6.4 Chapter summary

In this chapter we have investigated the robustness of the count of forbidden patterns as a criterion for the detection of determinism with respect to data that has been under-sampled, or irregularly sampled by a process of random depletion and timing jitter. For each flawed sampling scheme we generated
continuous model time series from data from the Lorenz system in both periodic and chaotic regimes, and the Rössler system in 8-periodic and hyper-chaotic regimes. From this data we computed the relative number of forbidden patterns for ordinal patterns up to $m = 8$. Our results demonstrated that the count of forbidden patterns is robust under certain conditions for all of the sampling schemes that were tested.

In the case of under-sampling for periodic time series we found that the count of forbidden patterns remains non-zero for significantly under-sampled data, and showed that it will remain non-zero for any regular sampling interval given sufficient $m$. For chaotic time series, successive time series points will rapidly de-correlate as the sampling interval increases due to the sensitivity to initial conditions. When the sampling interval is too large, the time series resembles a random process and, hence, forbidden patterns will no longer occur. We found it necessary for the sampling interval to be in the range where correlation can still clearly be observed from the autocorrelation function of the time series.

Where data has been randomly depleted from the time series we observe that the count of forbidden patterns will become unreliable when the length of data available approaches the number of possible permutations, as governed by the window length parameter $m$. For time series affected by timing jitter, our results indicate that the average density of sampling rather than the degree of jitter, is the primary factor that determines the reliability of the count of forbidden patterns.

However, the overarching finding in this investigation is that the window length $m$ should be chosen as large as possible for a given length $N$ of data available. Broadly speaking, our results agree with those of Kulp et al. [2016], yet we found that the count of forbidden patterns was significantly more robust when $m$ was chosen larger in all instances. Furthermore, we provided a simple analytical argument for the importance of selecting $m$ as large as possible for the case of random depletion.
Chapter 7

Summary and discussion

The subject of this thesis has been the definition and investigation of a new method of network based nonlinear time series analysis which we have termed ordinal network analysis. Ordinal networks are Markov chains constructed from a univariate time series by the application of an ordinal partition to a delay coordinate reconstruction of phase space. The purpose of this research has been to develop a network-based method of nonlinear time series analysis that is designed specifically to quantify information about the temporal evolution of a system. The fundamental hypothesis of this thesis is that the structural properties of an ordinal network bear a consistent and measurable relationship to the dynamical properties of time series from which the network was constructed. The scope of our investigation has been focussed on discretely sampled time series observed from archetypal continuous chaotic systems, and is based on the assumptions of stationarity and ergodicity. In summary, the contributions of this thesis pertain to: interpretation of the ordinal partition under the assumption of determinism; parameter selection; establishing relationships between simple measures of network structure and time series dynamics; developing and refining the novel measures respectively termed local node out-link entropy and inverse normalised weighted network diameter to quantify specific aspects of dynamical complexity; time series prediction and the generation of surrogates using ordinal networks; and the reliability of ordinal pattern statistics in irregularly sampled data. Key concepts, findings and contributions in these areas are discussed as follows.

An ordinal partition defines a set of $m!$ disjoint subspaces that span $m$-dimensional phase space. Applying the partition to a univariate time series requires a delay coordinate embedding. The topology of the partition with respect to the time series data is specified completely by the embedding dimension $m$ and embedding lag $\tau$. By increasing $m$, the embedded state vectors correspond to increasingly longer trajectories from the time series. In chaotic systems, nearby initial trajectories will diverge exponentially and thus eventually be mapped to distinct ordinal symbols once $m$ is sufficiently large. Therefore, $m$ effectively dictates the resolution of the partition over the data. We proposed that $m$ should be selected to minimise what we termed degeneracies in the network, which manifest when the partition does not distinguish between two or more state vectors that should not be considered as equivalent symbolic states for the purpose of the model. However, it is difficult to identify and measure the occurrence of network degeneracies without prior knowledge of the dynamics. We therefore proposed to select the embedding dimension based on maximising the local complexity of the ordinal network as quantified by mean out-degree or global node out-link entropy, and found that this produced
networks with structural properties that could be used to distinguish between periodic and chaotic
dynamics. Theoretically speaking, larger $m$ should produce an ordinal network that is a more accurate
model of the time series dynamics. What ultimately prohibits the selection of an arbitrarily fine
partition is the length of the time series $N$, which limits the extent to which we can ensure good
sampling over the partition. It is prudent to employ the standard heuristic that $N \gg m!$ (Bandt
and Pompe [2002]) when partitioning time series which may be stochastic or have a strong stochastic
component, thereby ensuring that all possible states are sampled. On the other hand, the total count
of distinct ordinal symbols observed in a deterministic time series generally grows with respect to $m$
at a much slower rate than $m!$ due to the phenomenon of forbidden patterns (Amigó et al. [2007]), hence,
we have noted that $N \gg m!$ may be an overly conservative criteria in some cases. As an alternative
criteria for sufficient sampling, we proposed to observe the growth of the number of nodes (equivalent
to the count of distinct ordinal symbols in the data) in the network constructed from a subset of
the data with length $k$ as we increase the length of that subset. If the network size saturates for
$k \ll N$ then we infer that the data is sufficiently long to ensure good sampling over the $m$-dimensional
ordinal partition, otherwise $N$ is likely too small for the choice of $m$. The fundamental purpose of
the $N \gg m!$ heuristic and our proposed criteria is to provide a basis from which it can be assumed
that the empirical probability distribution of the symbolic dynamics is a good approximation of the
true distribution. Therefore, we propose here that a more theoretically sound approach would be to
construct the ordinal symbolic dynamics for increasingly large subsets of the time series with length $k$,
and then quantify the rate of change of the empirical probability distribution using the total variation
distance between successive distributions. If the total variation distance converges within some small $\epsilon$
for $k \ll N$ then we might infer that the time series is long enough to ensure good sampling with respect
to the $m$-dimensional ordinal partition. If however, we are interested in the transitional distribution, it
would be necessary to observed the convergence of the conditional probability distribution or joint
probability distribution for symbol pairs. The appropriate heuristic to ensure sufficient sampling of the
conditional distribution in the general case would then be $m(m!) \ll N$.

We have argued in agreement with Groth [2005] that the embedding lag should be selected such
that the reconstructed attractor is unfolded over the ordinal partition in phase space. In the context
of ordinal analysis, an optimally unfolded attractor will occupy a maximal number of states over the
partition. We found that selecting embedding lag as the first zero of the time series autocorrelation
function resulted in ordinal networks which encoded useful information about the dynamics, and
proposed that other traditional heuristics, such as the first minimum of the time-lagged mutual
information, are also likely to be effective.

We identified a phenomenon which we termed node aliasing. This is a common artefact that can
occur in ordinal networks constructed from discrete-sampled time series where an edge skips over a
node in the network that would have been traversed by the continuous trajectory. This can occur if
the sampling frequency of the time series is too low relative to the rate of evolution of the trajectories
and the length span of the elements of the ordinal partition in embedding phase space in the direction
tangential to the flow. When node aliasing occurs, it falsely implies local uncertainty in the Markov
chain which can give rise to misleading results. Node aliasing can be reduced by selecting a smaller
embedding dimension to decrease the effective resolution of the partition. However, this will reduce
the amount of information encoded in the network. Alternatively, the time series can be interpolated. This is, of course, contingent on identifying an appropriate method of interpolation for the specific data in question, if one exists. We applied cubic spline interpolation on coarsely sampled data from the Rössler system to produce networks that were significantly more effective in distinguishing periodic dynamics from chaos. More importantly, the results for the interpolated time series were practically identical to those computed using highly sampled data from the same system.

We defined and investigated unweighted ordinal networks as simple stochastic models of non-linear time series. The number of nodes in an ordinal network can be interpreted as a measure of topological complexity. The value of this statistic relative to the total number of possible nodes quantifies the extent to which reconstructed phase space has been filled by the attractor. Node out-degree quantifies local complexity. Mean out-degree is an estimate of the global transitional complexity or uncertainty of the model, and the variance likely quantifies some aspect of the heterogeneity of reconstructed attractor. Our results show that these measures all exhibit sensitivity to changes in dynamics and track the relative change in estimates of the largest Lyapunov exponent along the bifurcation sequences of the Rössler system and an experimental diode resonator circuit. However, the fact that these simple network models are unweighted, creates a severe distortion of the probability distribution of states and the conditional or transitional probability distributions that are observed in the symbolic dynamics.

For this reason, we defined weighted ordinal networks where edge weights give the frequency of transitions between states. It is then trivial to generate a transition matrix that models the symbolic dynamics as a first order Markov chain. If the network is irreducible and aperiodic then the empirical probability distribution of the symbolic dynamics (the probability of observing a given state) corresponds to the stationary distribution of the Markov chain. We defined a new measure of transitional complexity called local node out-link entropy which is the Shannon entropy of the possible state transitions for a given node excluding self edges, and global node out-link entropy which is the network average of local node out-link entropy weighted by stationary distribution. Global node out-link entropy is very closely related to conditional permutation entropy proposed by Unakafov and Keller [2014] except that the latter permits the possibility of self edges. We show that in the limit of $\Delta t \to 0$, conditional permutation entropy converges to zero whereas global node out-link entropy remains effective in characterising dynamical complexity. Global node out-link entropy is bounded from below by zero, which implies periodic dynamics or a fixed point, and bounded from above by $\log(m! - 1)$. Numerical investigations show that global node out-link entropy tracks the relative change in the largest Lyapunov exponent more effectively than unweighted network measures. Results show that estimates of the largest exponent computed using the algorithm by Kantz [1994] are still significantly more reliable than global node out-link entropy, even though both measures essentially quantify the divergence of nearby states for chaotic time series. This is almost certainly due to the fact that Lyapunov exponent is being estimated over long trajectories that often span many cycles. Local node out-link entropy quantifies divergence over a single time step and will therefore be far more sensitive to noise and errors in the model such as network degeneracies and node aliasing. While it should be theoretically possible to obtain a more reliable measure by specifying large $m$ such that each ordinal symbol represents a long trajectory, it is impractical given that the length of data available in most applications will not permit accurate sampling of the dynamics for large $m$. 
A potential solution to these issues would be to compute conditional permutation entropy directly from the symbolic dynamics for symbols $s_n$ and $s_{n+\Delta n}$ and observe the growth of entropy with respect to $\Delta n$ in the spirit of the algorithms by Kantz [1994] and Rosenstein et al. [1993] for estimating the largest Lyapunov exponent.

Despite the noted issues, node entropy is a parametrically simple and computationally efficient measure. Taking into account previous studies relating to the relationships between embedding lag and ordinal analysis metrics, we designed a framework for multiscale ordinal network analysis whereby global node out-link entropy is computed for different time scales by varying the embedding lag. Results showed that multiscale global node out-link entropy is effective for discriminating between short time electrocardiogram recordings of normal sinus rhythm, ventricular tachycardia and ventricular fibrillation, and served to characterise differences in the transitional complexity of ordinal networks from interbeat interval time series on short and long time scales with respect to subject age.

The diameter of an unweighted ordinal network is a known lower bound for mixing time of a Markov chain if the network is irreducible and aperiodic (Levin et al. [2009]). We derived a new diameter bound for the mixing time based on the smallest non-zero probability of the stationary distribution. In addition, we showed that an ordinal network will always be irreducible if there exists a finite length path from the node corresponding to the final symbol in the symbolic dynamics, to the node corresponding to the first symbol. The assumption that a time series is both ergodic and stationary implies that this path will exist if the time series is sufficiently long. We found that the diameter of unweighted ordinal networks is highly sensitive to the interior crisis at the boundary between multiband chaos to chaos-chaos intermittency observed in the data from the diode resonator circuit. The path corresponding to the network diameter will exclude self edges by definition, which may lead to an excessively conservative estimate of mixing time. We proposed a method of re-weighting the network edges equal to the average time that the system spends in each respective state before transitioning to a different one by using additional information contained in the self edge weight of a weighted ordinal network. Specifically, we derive that the average time spent in a given state can be computed by dividing the self edge weight by the sum of the weights of all out-connected edges excluding the self edge. The weighted diameter of the modified network is therefore a better estimate of the mixing time.

We defined a new measure called the inverse normalised weighted network diameter. Results from numerical investigations show that this measure is closely related to the largest Lyapunov exponent for a range of low dimensional chaotic flows as it accurately tracks the relative change in the exponent along bifurcation sequences and reliably detects all periodic windows, even when the time series were affected by low levels of additive white Gaussian noise. We found that inverse normalised network diameter performs vastly better than global node out-link entropy, and for much smaller embedding dimensions, which is beneficial for sampling. This is likely because the path corresponding to the diameter represents a symbolic trajectory that can be generated by the model, and this trajectory is the maximum geodesic between any two states with respect to time. The inverse normalised network diameter is therefore measuring properties of the dynamics on an intermediate time scale.

We applied the inverse normalised weighted network diameter as a measure for classifying epileptic seizure onset in multivariate electroencephalogram time series. We performed a sliding window analysis to construct feature vectors comprising the scalar value of our new measure for each channel of the
record with respect to time. Binary classification models were constructed using a linear support vector
machine to show that the feature vectors corresponding to the pre-ictal phase were linearly separable
from those corresponding to the seizure such that the onset could be classified reliably using the new
measure. Our results also showed evidence of a system state change prior to the expertly annotated
onset time in some of the records. This investigation was performed in-sample as a proof of the concept
that inverse normalised network diameter has potential for application to clinical data sets. Given
the apparent preliminary success in this application, the next phase of this research should involve
either dividing the data into training and test data to correctly validate the model, or implementing
the measure within an established framework for change point detection.

Furthermore, we proposed a method for generating surrogate time series from ordinal networks and
investigated the extent to which the surrogates were similar to the original time series. We established
that treating an ordinal network as a first-order Markov chain and performing a random walk does not
guarantee that temporally overlapping ordinal symbols in the surrogates will have congruent order
patterns when the embedding lag is greater than one. To enforce topological congruency between
overlapping symbols we defined an algorithm for a constrained random walk. We employed this
algorithm to generate ordinal network surrogates for chaotic time series from the Lorenz and Rössler
systems and showed qualitative and quantitative similarities between the data and the surrogates.
Order recurrence plots and recurrence quantification analysis revealed relative consistency for both
vertical and diagonal line structures between the ordinal symbolic dynamics of the original data and
the symbol sequences generated by constrained random walk. However, measures of diagonal line
structures clearly diverge between the data and surrogates when the embedding dimension is such
that the number of nodes in the network approaches the length of the data. We have also found that
delay coordinate embeddings of the surrogate time series appear to exhibit exponential divergence
of trajectories, and that the estimated rate of this divergence is similar to that of the original data,
albeit over a shorter scaling region. This indicates that the ordinal network has the capacity to mimic
properties of chaotic dynamics, despite the model being stochastic. We then defined a method for time
series prediction using the ordinal network surrogates whereby we compute the expected value of an
ensemble of surrogates realisations which are all initialised at the state from which we want to make
the prediction. Numerical investigations showed that the ordinal network model was a significantly
better predictor than the mean of the training data for low dimensional continuous chaotic time series
generated by the Lorenz system. Predictions were generally accurate while trajectories remained in
one wing of the attractor but diverged rapidly over several cycles when the true trajectory passed
near or through the separatrix. The overarching implication is that the network model is encoding
significant information about the time series dynamics.

Finally, we investigated the reliability of the measure called the count of forbidden patterns when
time series are irregularly sampled by processes of random depletion or timing jitter. Forbidden patterns
correspond to elements of the ordinal partition of reconstructed phase space that embedded trajectories
do not visit. The count of forbidden patterns can, therefore, be interpreted as a measure of topological
complexity. It is directly related to the number of nodes in an ordinal network as its compliment
with respect to the complete set of possible symbols for a given embedding dimension. Moreover, we
have discussed the well established result by Amigó et al. [2007] that a non-zero count of forbidden
patterns theoretically implies determinism in the limit of infinitely long data. We performed a series of numerical experiments using irregularly sampled time series from the Lorenz and hyper-chaotic Rössler system, respectively. Irregular sampling adds a stochastic element to the time series and leads to a reduction of the count of forbidden patterns, as shown by our results. However, the key finding is that the measure is robust, in that it remains non-zero, against severe random depletion of data and timing jitter when the embedding dimension is selected sufficiently large. This finding is dependent on the condition that the time series comprises significantly more observations than the total number of possible ordinal symbols to ensure accurate sampling. In addition, we demonstrated via a simple analytical argument that a larger embedding dimension will always result in increased reliability against random depletion where sufficient data is available. A non-zero count of forbidden patterns cannot be used as a criteria for determinism in practice due to finiteness of data. However, our findings indicate that existing statistical methods which adapt this measure as a tool for detecting determinism is noisy finite time series (Amigó et al. [2010]) may also be effective for irregularly sampled data and, therefore, warrant further investigation.
Bibliography


